

Aminosäurepolymorphismen bei funktionell
verschiedenen Allelen des geschlechtsbestimmenden
Gens *complementary sex determiner (csd)* bei der
Honigbiene

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Christine Seelmann

aus Worms

Düsseldorf, April 2011

aus dem Institut für Genetik
der Heinrich-Heine Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Martin Beye
Koreferent: Prof. Dr. Klaus Lunau

Tag der mündlichen Prüfung: 10.06.2011

Danksagung

Herrn Professor Martin Beye möchte ich für die Überlassung des Themas und die wissenschaftliche Betreuung danken.

Sehr herzlich möchte ich mich bei Miriam Müller und Björn Schmitt für intensive Korrekturarbeiten bedanken.

Ebenso danke ich Tanja Gempe und Martin Hasselmann für das Korrekturlesen von Teilen dieser Arbeit.

Frau Didina David danke ich sehr für die Hilfe bei Arbeiten im Labor und bei technischen Problemen.

Mein Dank gilt allen Mitgliedern der Arbeitsgruppe von Professor Beye für anregende Diskussionen und nützliche Ratschläge und Ideen.

Meinem Bruder Albrecht danke ich für die Hilfe bei mathematischen Problemen.

Schließlich möchte ich meinen Eltern für ihre ständige Unterstützung danken.

I. Einleitung	1
I.1 Geschlechtsbestimmungen im Tierreich laufen nach einem grundsätzlichen Prinzip ab.....	1
I.2 Der SDL und die Geschlechtsbestimmung bei <i>Apis mellifera</i>	3
I.3 Hypothetische Modelle zur Proteinaktivierung.....	5
I.4 Der Einfluss balancierender Selektion auf das <i>csd</i> -Gen.....	8
I.5 Zielsetzung der Arbeit.....	10
II. Material und Methoden	11
II.1 Chemikalien und Kitsysteme.....	11
II.2 DNA-Extraktion.....	12
II.2.1 DNA-Extraktion aus Drohnen und Arbeiterinnen (modifiziert nach Hunt & Page 1995).....	12
II.2.2 DNA-Extraktion aus Einzeleiern.....	13
II.3 Total-RNA-Präparation aus Eierpools und cDNA-Synthese.....	13
II.4 PCR-Reaktionen und Restriction-Fragment-Length-Polymorphism (RFLP)-Analyse.....	15
II.5 Klonierungen.....	21
II.6 Experimenteller Aufbau.....	23
II.6.1 Kreuzungsexperiment.....	23
II.6.2 Bestimmung der Consensussequenzen der klonierten Nukleotidsequenzen mit Hilfe des Programmes GAP (Staden Package).....	27
II.6.3 Zusammensetzen der <i>csd</i> -Sequenzen aus den Klonierungen der 5' UTR bis Exon 6/7/8 Regionen und der Exon 6-9 Region.....	27
II.6.4 Untersuchung der phylogenetischen Verwandtschaft der <i>csd</i> -Allele mit Hilfe des Programmes Molecular Evolutionary Genetics Analysis (MEGA)....	28
II.7 Erstellung von Datenmatrices für die Analyse mit Zwei-Weg hierarchischer Clusteranalyse und Multidimensionaler Skalierung.....	29
II.7.1 Erstellen einer Datenmatrix basierend auf Aminosäureunterschieden: Matrix 1.....	29
II.7.2 Erstellen einer Datenmatrices basierend auf physikochemischen Differenzen von substituierten Aminosäuren: Matrix 2 und Matrix 3.....	31
II.8 Datenanalyse mit Hilfe von Zwei-Weg hierarchischer Clusteranalyse und Multidimensionaler Skalierung.....	36
II.8.1 Transformation der Datenmatrices in Matrices, die die paarweise Ähnlichkeit der Daten darstellen (Ähnlichkeitsmatrices.....	36
II.8.1.1 Berechnung einer Ähnlichkeitsmatrix mittels Spearman'scher Rangkorrelation.....	36
II.8.1.2 Berechnung einer Ähnlichkeitsmatrix mittels Bildung der Euklidischen Distanzen zwischen den Datenreihen.....	38
II.8.2 Analyse der (Un)ähnlichkeitsmatrices mittels Zwei-Weg hierarchischer Clusteranalyse und Multidimensionaler Skalierung.....	39
II.8.2.1 Zwei-Weg hierarchische Clusteranalyse mit Hilfe des Programms „Cluster 3.0“.....	39
II.8.2.2 Multidimensionale Skalierung (MDS) mit Hilfe des Programms „Primer 6“..	40

III. Ergebnisse	41
III.1 Bestimmung von Nukleotidsequenzen und funktionellen Unterschieden der im Kreuzungsexperiment kombinierten <i>csd</i> -Allele.....	41
III.1.1 Das Kreuzungsexperiment.....	41
III.1.2 Identifikation der <i>csd</i> -Genotypen der P- und F1- Generationen und der F2-Drohnen.....	43
III.1.3 Identifikation funktionell verschiedener <i>csd</i> -Allele in den Arbeiterinnen der F2-Generation.....	46
III.1.4 Bestimmung der Aminosäuresequenzen <i>csd</i> -Allele.....	51
III.2 Analysen zur phylogenetischen Verwandtschaft der 14 <i>csd</i> -Allele.....	52
III.3 Bearbeitung der Sequenzdaten für die Analyse mit Hilfe von Zwei-Weg hierarchischer Clusteranalyse und Multidimensionaler Skalierung.....	54
III.3.1 Transformation der Daten aus den Sequenzvergleichen in Datenmatrices...	54
III.3.1.1 Datenmatrix basierend auf Aminosäureunterschieden: Matrix 1.....	54
III.3.1.2 Datenmatrix basierend auf physikochemischen Differenzen der Aminosäureunterschiede: Matrix 2.....	56
III.3.1.3 Datenmatrix basierend auf physikochemischen Aminosäureunterschieden, unter Ausschluss der Aminosäurepositionen, die Deletionen aufweisen: Matrix 3.....	57
III.3.2 Bestimmung der Ähnlichkeit von Datenreihen durch Erstellen von Ähnlichkeits- bzw. Unähnlichkeitsmatrices.....	57
III.4 Analysen der (Un)Ähnlichkeitsmatrices mittels Zwei-Weg Hierarchischer Clusteranalyse und Multidimensionaler Skalierung.....	58
III.4.1 Analysen der Matrix 1.....	59
III.4.1.1 Zwei- Weg Hierarchische Clusteranalyse von Matrix 1.....	59
III.4.1.1.1 Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher Rangkorrelation.....	59
III.4.1.1.2 Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen.....	65
III.4.1.2 Multidimensionale Skalierung (MDS) der Allele von Matrix 1.....	67
III.4.1.2.1 Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher Rangkorrelation.....	67
III.4.1.2.2 Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen.....	68
III.4.1.3 Multidimensionale Skalierung der Aminosäurepositionen von Matrix 1.....	70
III.4.1.3.1 Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher Rangkorrelation.....	70
III.4.1.3.2 Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen.....	70
III.4.2 Zwei-Weg Hierarchische Clusteranalyse von Matrix 2.....	73
III.4.2.1 Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher Rangkorrelation.....	73
III.4.2.2 Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen.....	79
III.4.3 Analysen der Matrix 3.....	86
III.4.3.1 Multidimensionale Skalierung der Allele von Matrix 3.....	86
III.4.3.1.1 Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher	

	Rangkorrelation.....	86
III.4.3.1.2	Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen.....	86
III.4.3.2	Multidimensionale Skalierung der Aminosäurepositionen von Matrix 3	88
III.4.3.2.1	Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher Rangkorrelation.....	88
III.4.3.2.2	Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen.....	88
III.5	Separate Analyse der Aminosäurepositionen, die Deletionen aufweisen.....	92
III.6	Vergleich der Ergebnisse der Zwei-Weg hierarchischen Clusteranalyse und der Multidimensionalen Skalierung.....	93
III.6.1	Vergleich der Ergebnisse der Allelanalysen.....	93
III.6.2	Vergleich der Ergebnisse der Analysen der Aminosäurepositionen.....	95
III.7	Separate Auswertung der hypervariablen Region (HVR).....	103
III.8	Vergleich der Typeinteilung der hypervariablen Regionen mit den auf den ZWHC und den MDS basierenden Ähnlichkeitsdarstellungen der <i>csd</i> -Allele.	105
III.9	Ausführlicher Vergleich der Allele G1, G2 und Y2.....	106
 IV. Diskussion		 109
IV.1	Die experimentell kombinierten <i>csd</i> -Allele sind funktionell verschieden.....	109
IV.2	Außer bei den Allelen G1 und Y2 konnte keine phylogenetische Verwandtschaft der <i>csd</i> -Allele nachgewiesen werden.....	110
IV.3	Welche Regionen des <i>csd</i> -Gens könnten für die Spezifität der Allele besonders wichtig sein?.....	110
IV.4	Die Ergebnisse der Zwei-Weg hierarchischen Clusteranalyse und der Multidimensionalen Skalierung können in Abhängigkeit von unterschiedlichen Hypothesen interpretiert werden.....	113
IV.5	Welche Aminosäurepositionen könnten potentiell wichtig für die Spezifität von <i>csd</i> -Allelen sein?.....	116
IV.6	Wie beeinflussen die physikochemischen Differenzen der Aminosäuresubstitutionen und die Aminosäuredeletionen die Ähnlichkeitsdarstellungen der Allele?.....	120
IV.7	Aminosäuredeletionen könnten die Spezifität von <i>csd</i> -Allelen beeinflussen	123
IV.8	Es wurden Hinweise auf Koevolution von Aminosäurepositionen gefunden	125
IV.9	Die hypervariable Region.....	129
IV.10	Die HVR-Typen sind nicht mit bestimmten Mustern von Aminosäuresubstitutionen in übrigen Teil des <i>csd</i> -Gens assoziiert.....	132
IV.11	Die Spezifität von <i>csd</i> -Allelen ist nicht von einzigartigen Aminosäuresubstitutionen abhängig.....	133
IV.12	Die Spezifität von <i>csd</i> -Allelen könnte über spezifische Bindungseigenschaften der RS-Domäne und der prolinreichen Region kodiert werden.....	134
IV.13	Die Allele G1, G2 und Y2 zeigen Spuren eines Rekombinationsereignisses...	136
IV.14	Diskussion der Methoden.....	139
IV.15	Fehlerdiskussion.....	141

IV.16	<i>In vivo</i> Versuche zum Testen bestimmter Genregionen von <i>csd</i> auf ihre Bedeutung für die Spezifität von <i>csd</i> -Allelen.....	143
	Zusammenfassung.....	145
	Abstract.....	147
	Literatur.....	148
	Anhang.....	A.1

I. Einleitung

I.1 Geschlechtsbestimmungen im Tierreich laufen nach einem grundsätzlichen Prinzip ab

Im Tierreich sind sehr unterschiedliche Arten der Geschlechtsbestimmung verbreitet. Das Geschlecht kann z.B. durch bestimmte Umweltfaktoren festgelegt werden (modifikatorischen Geschlechtsbestimmung) oder mittels Geschlechtschromosomen. Ein weiterer Mechanismus der Geschlechtsdetermination ist die Haplodiploidie. Obwohl das initiale Signal zwischen diesen geschlechtsbestimmenden Mechanismen variiert, gibt es ein grundsätzliches Prinzip. Nachdem die primäre Entscheidung gefallen ist, wird eine Genkaskade initiiert, im Laufe derer einer von zwei alternativen Reaktionswegen eingeschlagen wird. Dies führt dann zur Ausdifferenzierung männlicher oder weiblicher Geschlechtsmerkmale. Beim Alligatoren ist das initiale Signal zum Beispiel die Bruttemperatur. Bei vielen Säugetierarten und dem Mensch wird das Geschlecht in Abhängigkeit von Geschlechtschromosomen bestimmt. Dabei dient ein einziger dominanter männlicher Faktor (*Sry*) auf dem Y-Chromosom als initiales Signal (Goodfellow and Lovell-Badge, 1993). Bei *Drosophila melanogaster* und *Caenorhabditis elegans* wird das Verhältnis von Gonosomen zu Autosomen (X:A- Verhältnis) und ein polygenisches Signal zur Geschlechtsbestimmung genutzt. Im XX-Genotyp *sisterless (sis):: daughterless (da)* aktivieren Heterodimere den frühen Promotor P_e des Gens *Sex-lethal (Sxl)* und so wird der weibliche Entwicklungsweg eingeleitet (Cline and Meyer, 1996). Bei 20% aller Tierarten (u.a. die gesamte Insektenordnung der Hymenopteren) wird das Geschlecht durch den Mechanismus der Haplodiploidie festgelegt (Bull, 1983; Bell, 1982). Diese Art der Geschlechtsbestimmung wurde erstmals von Whiting (1933 und 1943) in der parasitisch lebenden Wespe *Bracon hebetor* beschrieben. Die Entscheidung über das Geschlecht basiert bei der Haplodiploidie nicht auf Geschlechtschromosomen, sondern darauf, ob ein Ei befruchtet wurde oder nicht. Unbefruchtete (haploide) Eier entwickeln sich zu Männchen und befruchtete (diploide) Eier zu Weibchen (Bull, 1983). Die genetische Basis der Geschlechtsdetermination bei haplodiploiden Organismen ist noch weitgehend ungeklärt. Näher untersucht wurde sie bereits bei einigen Hymenopteren, vor allem bei Bienen und der Wespe *Bracon hebetor*. Whiting fand bei *Bracon hebetor* diploide Männchen und schloss daraus, dass weder die Befruchtung noch der diploide Zustand für die Geschlechtsbestimmung verantwortlich sein

können. Dies führte zu der Hypothese der komplementären Geschlechtsbestimmung, einer Variante der haplodiploiden Geschlechtsbestimmung, die von Whiting (1943) formuliert wurde. Bei diesem Mechanismus der Geschlechtsdetermination entscheidet der allelische Zustand am geschlechtsbestimmenden Locus (SDL) über das Geschlecht. Individuen, die am SDL heterozygot sind, entwickeln sich zu Weibchen, homozygote oder hemizygote (haploide) zu Männchen (Abbildung I.1). Homozygote Männchen sind entweder steril oder produzieren sterile, triploide Nachkommenschaft (Oishi et al., 1993; Cook, 1993). Im Falle der Honigbiene (*Apis mellifera*) werden diploide Männchen kurz nach dem Schlupf von Arbeiterinnen gefressen (Mackensen, 1955).

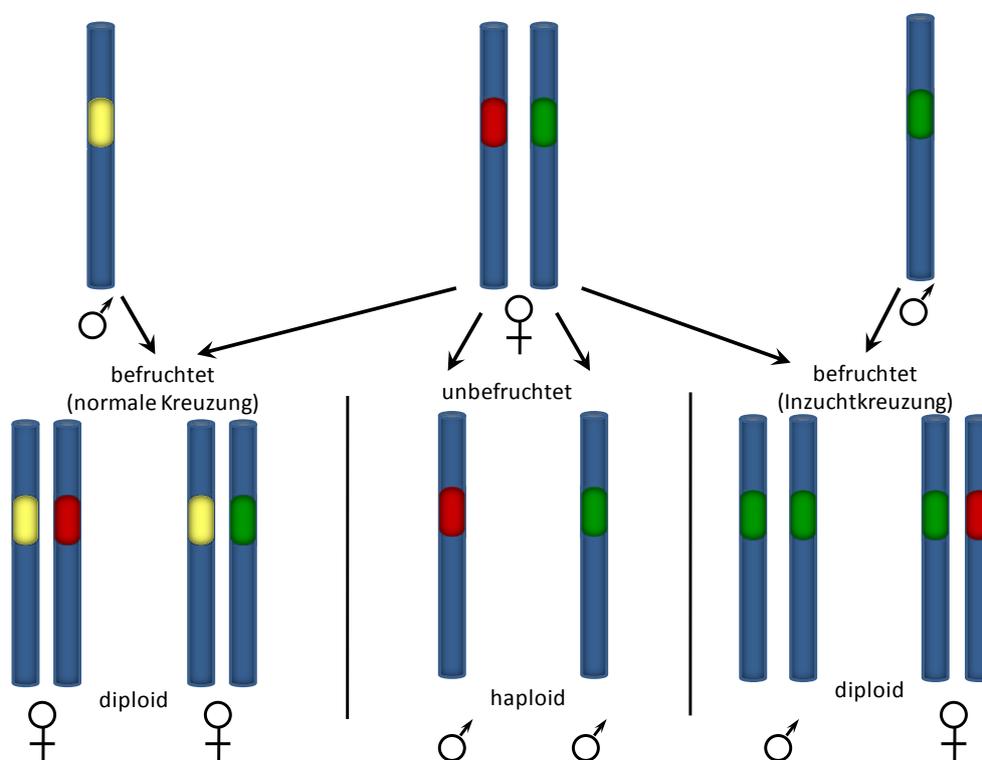


Abbildung I.1: Komplementäre Geschlechtsbestimmung bei Hymenopteren.

Dargestellt sind eine normale Kreuzung (linke Seite) und eine Inzuchtkreuzung (rechte Seite). Bei der normalen Kreuzung wird das Weibchen von einem Männchen besamt, welches keines ihrer geschlechtsbestimmenden Allele trägt (dargestellt durch unterschiedliche Farbe). Die diploide Nachkommenschaft ist dann zu 100 % heterozygot am SDL und entwickelt sich zu weiblichen Individuen. Bei der Inzuchtkreuzung trägt das inseminierende Männchen eines der Allele des Weibchens (grün dargestellt). Die Nachkommenschaft ist dann zu 50 % heterozygot und zu 50 % homozygot am SDL. Aus den Homozygoten entstehen diploide Männchen. Unbefruchtete Eier, dargestellt in der Mitte, entwickeln sich zu haploiden Männchen.

I.2 Der SDL und die Geschlechtsbestimmung bei *Apis mellifera*

Der SDL von *Apis mellifera* wurde durch M. Beye und Mitarbeiter identifiziert (Hasselmann et al., 2001; Hasselmann and Beye, 2003). Dafür wurde die Region, die bei weiblichen Bienen immer heterozygot vorliegt, mittels genetischer Marker eingegrenzt. Der SDL der Honigbiene umfasst fünf Gene, von denen zwei, nämlich das *complementary sex determiner (csd)* Gen und das *feminizer (fem)* Gen nachweislich in die Geschlechtsbestimmung involviert sind (Hasselmann et al. 2008 (2), Gempe et al. 2009). *csd*, welches durch Duplikation aus *fem* entstanden ist (Hasselmann et al. 2008 (2)), liegt Schätzungen zufolge in Form von 11 bis 19 verschiedenen Allelen in einer Population vor (Mackensen, 1955, Laidlaw et al. 1956; Adams et al. 1977). *fem* dagegen ist weitgehend konserviert mit nur geringer Sequenzvariation (Hasselmann and Beye, 2004, *fem* entspricht hier den Typ II - Sequenzen). Der allelische Zustand am *csd*-Lokus ist das primäre Signal, welches die Geschlechtsdifferenzierung einleitet (Beye et al., 2003, Hasselmann et al., 2008; Gempe et al., 2009). *csd* liegt bei weiblichen Bienen immer heterozygot vor (Beye and Hasselmann, 2003). Für *csd* hemi-oder homozygote Individuen entwickeln sich zu männlichen Bienen (Drohnen). Homozygote Männchen entstehen vor allem bei Inzuchtkreuzungen, wenn die Königin mit den Spermien ihrer Geschwisterdrohnen besamt wird. Die verschiedenen *csd*-Allele unterscheiden sich im Durchschnitt zu 3 % in ihrer Aminosäuresequenz (Hasselmann and Beye, 2004; Hasselmann et al., 2008 (1)). Es wird davon ausgegangen, dass nicht alle allelischen Unterschiede in der Aminosäuresequenz zur Spezifität von *csd*-Allelen beitragen. Allele, die sich nur geringfügig in ihrer Aminosäuresequenz unterscheiden, sind vermutlich funktionell gleiche Replikatvarianten einer Allellinie (Hasselmann and Beye, 2004). Daher ist ein Individuum nicht nur homozygot, wenn es zwei gleiche *csd*-Sequenzen, sondern auch, wenn es zwei Replikatvarianten derselben Allellinie trägt. Nur Individuen, die zwei funktionell verschiedene *csd*-Allele tragen, sind für *csd* heterozygot.

Der Open Reading Frame der aus neun Exonen bestehenden *csd*-Sequenzen reicht von der Mitte des Exon 2 bis zum Anfang des Exon 9 (Gempe et al., 2009) und kodiert für ein Protein mit einer RS (Arginin-Serin)-reichen Domäne und einer prolinreichen Region am C-Terminus (Beye et al., 2003). Zwischen diesen beiden Domänen liegt die hypervariable Region (HVR). Abbildung I.2 zeigt die Schematische Darstellung der Domänen und Regionen des Csd-Proteins und der entsprechenden Exone, die die Domänen und Regionen kodieren. Die HVR

besteht hauptsächlich aus den Aminosäuren Asparagin (N) und Tyrosin (Y). Diese sind in Repeats angeordnet, in denen auf mehrere Asparaginreste ein Tyrosin folgt. Die Anzahl dieser Repeats kann zwischen den verschiedenen Allelen stark variieren, so dass Längendifferenzen zwischen Allelen auftreten (Beye, 2004).



modifiziert nach Gempe, 2006

Abbildung 1.2: Schematische Darstellung der Abfolge der verschiedenen Domänen / Regionen des Csd-Proteins und den entsprechenden Exonen des *csd*-Gens. In blau ist die N-terminale Region des Proteins dargestellt, in grün die RS-Domäne (RS), in gelb die hypervariableRegion (HVR) und in rot die c-terminal gelegene prolinreiche Region (P). Die Exone 2 bis 9 (E 2 bis 9) sind als Pfeile dargestellt.

Es wurden keine geschlechtsspezifischen *csd*-Allele und auch keine geschlechtsspezifischen Spleißformen der *csd*-Transkripte gefunden (Beye et al., 2003). Das Gen *fem*, welches „downstream“ von *csd* in der geschlechtsdeterminierenden Kaskade lokalisiert ist, besitzt geschlechtsspezifische Spleißformen (Hasselmann et al., 2008 (2); Gempe et al., 2009). Nur das weiblich gespleißte *fem* Transkript kodiert für ein aktives Protein. Dieses FEM Protein reguliert das Spleißen von Transkripten des Gens *Am-doublesex* (*Amdsx*) in dessen weibliche Form (Gempe et al., 2009). *Amdsx* ist ein weiteres Mitglied der geschlechtsbestimmenden Kaskade von *Apis*, und sorgt so dafür, dass der weibliche Differenzierungsweg eingeschlagen wird.

Mit RNAi–knockdown Versuchen wurde gezeigt, dass das Genprodukt von *csd*, das Csd-Protein, nur für die weibliche, nicht aber für die männliche Entwicklung benötigt wird (Beye et al., 2003; Hasselmann et al., 2008 (2); Gempe et al., 2009). Das bedeutet, dass Csd nur in heterozygotem Zustand aktiv ist. Der hemi- oder homozygote männliche Zustand ist inaktiv (Gempe et al., 2009). Als Reaktion auf den heterozygoten Zustand von Csd wird *fem* mRNA geschlechtsspezifisch prozessiert. Dies lässt vermuten, dass *fem* direkt vom Aktivitätszustand von Csd beeinflusst wird und vermutlich dessen Zielgen ist (Gempe et al., 2009). Die

Aktivierung von heterozygotem Csd zeigt analoge Strukturen zum Selbstinkompatibilitätssystem in *Arabidopsis thaliana* (Nasrallah, 2002). Der S-Lokus von *A. thaliana*, der in verschiedenen Allelen in einer Population vorliegt, enthält zwei Gene: das SRK (S-Lokus Rezeptor Kinase) Gen und das SCR (S-Lokus cysteinreiches Protein) Gen. Die Proteine, die von diesen Genen kodiert werden, sind die Determinanten der Selbstinkompatibilität von Stigma und Pollen. Bei der Bestäubung registrieren Zellen der Stigma-Epidermis Pollen, die ein S-Lokus Allel besitzen, das einem ihrer Allele funktionell gleich ist und inhibieren die Befruchtung durch diesen Pollen.

I.3 Hypothetische Modelle zur Proteinaktivierung

Das *csd*-Gen kodiert für ein SR-(Serin-Arginin) Typ Protein (Hasselmann et al. 2008 (2)). Es wurde gezeigt, dass Proteine mit RS-Domänen proteinbindende Funktion haben können (Graveley, 2000; Tacke and Manley, 1999; Kay et al. 2000) Mattox and Dauwalder (1998) haben mit Hilfe von SWAP Experimenten herausgefunden, dass RS-Domänen unterschiedliche Bindungsspezifitäten besitzen können. Außerdem wird vermutet, dass SR-Typ Proteine in die Regulation von RNA-Spleißvorgängen involviert und somit potentielle Spleißfaktoren sind (Blencowe et al., 1999; Tanabe et al., 2009; Cazallah et al., 2005).

Das Csd-Protein besitzt keine RNA-Erkennungssequenz (RRM). Wir vermuten daher, dass ein in der Honigbiene existierender Faktor mit RNA-bindender Funktion, kooperativ mit Csd an die pre- mRNA seines Zielgens *fem* bindet und so das Spleißen dieser pre- mRNA vermittelt (Beye et al., 2003).

In natürlichen Populationen segregieren bis zu 19 *csd*-Allele (Adams et al. 1977). Dies ermöglicht bis zu 171 heterozygote Allelkombinationen, in denen das Csd-Protein aktiviert ist. In bis zu 19 Kombinationen sind die Allele homozygot und das Csd-Protein inaktiv (Beye 2004). Aber wie wird das polymorphe Signal aus vielen funktionell verschiedenen Allelen in das binäre Signal, das zu männlicher oder weiblicher Entwicklung führt, verwandelt? Bisherige Modelle basieren auf aktiven heteromeren RNA- oder Protein- Molekülen, die bei heterozygotem Zustand gebildet werden (Crozier, 1971; Cook, 1993). Beye (2004) hat drei mögliche Modelle zur Assoziation zweier funktionell verschiedener Csd-Proteine vorgeschlagen (Abbildung I.3). Der Unterschied dieser drei Modelle besteht darin, welche Bindungsfähigkeiten den Csd-Proteinen zugeschrieben werden. Bei Modell 1 wird der aktive Proteinzustand dadurch

erreicht, dass nur Proteine sich verbinden, die von zwei funktionell verschiedenen *csd*-Allelen kodiert werden und so ein Heterodimer bilden. Dieses stellt dann den aktiven Proteinzustand dar, der nur in den heterozygoten Weibchen zu finden ist. Da Männchen homo-oder hemizygot sind und daher nur ein *csd*-Allel tragen, können sich bei ihnen keine Heterodimere bilden. Es existieren nur inaktive Monomere. Ist das Protein inaktiv, wird der voreingestellte, männliche Entwicklungsweg eingeschlagen. Dieses Modell ist dem Selbstinkompatibilitätsmechanismus des Pilzes *Ustilago* analog (Kamper et al., 1995). Im zweiten Modell ist das Bindungsverhalten der Csd-Proteine ebenfalls abhängig davon, ob zwei Proteine von dem gleichen Allel oder von zwei verschiedenen Allelen stammen. Hier können aber, im Gegensatz zu Modell 1, nur Proteine des gleichen Allels aneinander binden. Homodimere existieren dann in beiden Geschlechtern. Der Unterschied zu Modell 1 besteht darin, dass es bei Weibchen zwei verschiedene Arten von Homodimeren gibt, während Männchen nur eine Form von Homodimer besitzen. Ein solcher Mechanismus würde ein System erforderlich machen, das es ermöglicht, die Dosisunterschiede zwischen homozygoten und heterozygoten Allelkombinationen zu messen. Bei diesem Modell ist es also nicht das Csd-Protein, welches nachfolgende, geschlechtsdeterminierende Schritte ausführt, sondern ein unbekanntes System, das die Dosis von verschiedenen Homodimeren zueinander misst und diese Information in ein Signal verwandelt, welches dann weitere geschlechtsdeterminierende Mechanismen in Gang setzt. Die Grundidee von Modell 3 ist die funktionelle Komplementation. Es existieren keine Bindungsunterschiede, sondern Aktivitätsunterschiede zwischen den Csd-Proteinen, die von gleichen Allelen und die von verschiedenen Allelen stammen. Stammen alle Csd-Proteine eines Individuums nur von einem Allel, so sind diese Proteine inaktiv, unabhängig davon, ob sie als Monomer oder als Homodimer vorliegen. Nur die Kombination von funktionell unterschiedlichen Allelen im heterozygoten Zustand resultiert in einer funktionellen Komplementation und bringt aktive Proteine hervor (zwei verschiedene Monomere oder Heterodimere). Bei diesem Modell könnten es die funktionell verschiedenen Polypeptide sein, die eine Maschinerie anstoßen, die die geschlechtsbestimmende Kaskade in Gang setzt oder ein System, das die funktionelle Komplementation messen kann.

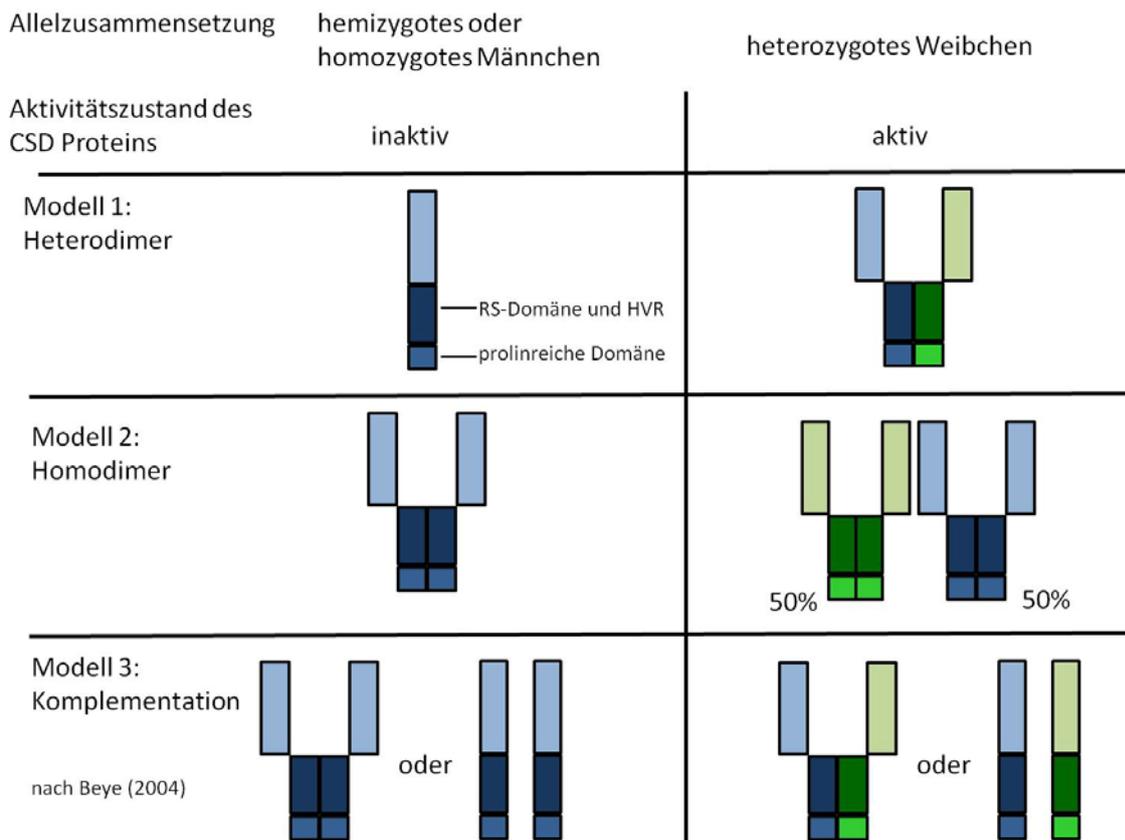


Abbildung 1.3: Modelle für die Assoziation zweier allelischer Csd-Proteine. Dargestellt sind drei Modelle, in denen die vorgeschlagene Bindungseigenschaft und die funktionelle Konsequenz für das Csd-Protein (aktiv bzw. nichtaktiv) angegeben sind. Die Regionen größter Variabilität des *csd*-Gens sind dunkelblau bzw. dunkelgrün (RS-Domäne, hypervariable Region) und mittelblau bzw. mittelgrün (prolinreiche Region) markiert. Der stärker konservierte N-terminale Bereich ist in hellblau bzw. hellgrün dargestellt.

Da die hypervariable Region von *csd* bei den verschiedenen Allelen sehr divers ist, liegt die Vermutung nahe, dass sie bei dem Zustandekommen von Spezifität von *csd*-Allelen eine entscheidende Rolle spielt (Hasselmann and Beye, 2004). Die beobachteten Längendifferenzen führen dazu, dass auch durch das Fehlen von Aminosäuren Volumendifferenzen zwischen den hypervariablen Regionen der verschiedenen *csd*-Allele existieren. Volumenunterschiede könnten die Tertiärstruktur im Bereich der hypervariablen Region ändern und benachbarte Domänen könnten davon beeinflusst werden. Eventuell hat auch die Zusammensetzung der hypervariablen Region selbst funktionelle Bedeutung (Beye, 2004). Auch unterschiedliche Phosphorylierungsstellen könnten möglicherweise für Spezifität verantwortlich sein. Es wurde gezeigt, dass die Funktion von SR-Proteinen durch Phosphorylierung ihrer RS-Domänen durch verschiedene Kinasen reguliert werden kann (Wang et al., 1998; Yeakley et al., 1999, Xiao and Manley 1998). Schließlich sind auch

Kombinationen aus all diesen Möglichkeiten zu erwägen.

I.4 Der Einfluss balancierender Selektion auf das *csd*-Gen

Das *csd*-Gen ist Ziel balancierender Selektion (Hasselmann et al, 2008 (1); Charlesworth, 2004; Hasselmann and Beye, 2004) und erhält dabei viele funktionell verschiedene Allele aufrecht (Hasselmann et al., 2008 (1); Hasselmann and Beye, 2004; Beye et al., 2003). Wie schon erwähnt, werden homozygote Männchen der Honigbiene direkt nach dem Schlupf von Arbeiterinnen gefressen, so dass deren Allele nicht in die nächste Generation übertragen werden. Es herrscht also eine starke Selektion gegen Homozygote. Dies führt dazu, dass die Selektion von *csd* Allelen frequenzabhängig ist (Hasselmann et al., 2008 (1)). Bei seltenen oder durch Mutation neu entstandenen Allelen ist die Wahrscheinlichkeit für die Entstehung von Homozygoten sehr gering. Je öfter ein Allel vorhanden ist, desto größer ist die Wahrscheinlichkeit, dass Individuen entstehen, die für dieses Allel homozygot sind. Davon profitieren seltene Allele, da sie oft in Heterozygoten vereint werden und somit ihre Frequenz in der Population zunimmt. Häufige Allele haben einen selektiven Nachteil, weil sie in den natürlichen Populationen oft in nicht überlebenden Homozygoten vereint werden (Kimura and Crow, 1964; Yokohama and Nei, 1979). Ihre Frequenz in der Population nimmt daher ab. Der Vorteil seltener Allele wird auch als negative Frequenzabhängigkeit oder balancierende Selektion bezeichnet. Allele, die unter balancierender Selektion stehen, gehen selten durch genetische Drift verloren und haben so eine längere Verweildauer in Populationen, als neutrale Allele (Takahata, 1990). Regionen, die unter balancierender Selektion stehen, können mit der Zeit deutlich mehr Nukleotidsubstitutionen akkumulieren, als Regionen, die nicht unter balancierender Selektion stehen und auch nicht mit Regionen gekoppelt sind, die unter balancierender Selektion stehen (Hasselmann et al., 2008 (1)).

Einige stark divergente Sequenzen von verschiedenen *csd*-Allellinien sind Konsequenz des hohen Alters dieser Allele (Hasselmann and Beye, 2004). Das *csd*-Gen hat im Durchschnitt 10-13-mal mehr Nukleotidpolymorphismen (Nukleotidsubstitutionen) akkumuliert als der genomweite Durchschnitt (Hasselmann and Beye, 2006). Bei Nukleotidpolymorphismen lassen sich synonyme und nichtsynonyme Substitutionen unterscheiden. Bei synonymen Substitutionen wird die Identität der kodierten Aminosäure nicht verändert, so dass sie

keinen Einfluss auf das Protein haben und daher nicht unter Selektion stehen. Zur Beurteilung, ob eine Region unter balancierender Selektion steht, werden die synonymen Polymorphismen dieser Region analysiert, denn die synonymen Polymorphismen werden unbeeinflusst von Selektion mit steigender Verweildauer der Allele akkumuliert, so dass der Einfluss der balancierenden Selektion an ihnen deutlich zu erkennen ist. Im Falle der nichtsynonymen Substitutionen wird die kodierte Aminosäure geändert. Dies kann Auswirkungen auf die Tertiärstruktur des resultierenden Proteins haben. Nichtsynonyme Substitutionen sind daher Angriffspunkte für negative Selektion. Durch negative Selektion werden die Mutationen entfernt, die die Proteinfunktion negativ beeinflussen. Die nichtsynonymen und synonymen Polymorphismen sind bei den verschiedenen Exonen von *csd* sehr heterogen verteilt, mit einem deutlichen Maximum beider Polymorphismenformen in den Exonen 6 und 7 (Hasselmann and Beye, 2006, Hasselmann et al., 2008 (1)). Der Nukleotidpolymorphismus nimmt mit Abstand zu den Exonen 6 und 7 ab (Hasselmann and Beye, 2006; Hasselmann et al., 2008 (1)). Es wird daher angenommen, dass die Exone 6 und 7 mit ihrem Maximum an synonymen Polymorphismen das Ziel der balancierenden Selektion sind und die potentiell pezifizierende Domäne (PSD) von Csd kodieren (Hasselmann et al., 2008 (1)). Die Akkumulation von nichtsynonymen und „stillen“ synonymen Nukleotiddifferenzen in den Exonen 2 bis 5 und 8 könnte abhängig sein von der Stärke der Kopplung dieser Exone mit den Exonen 6 und 7, denn in Regionen, die mit Zielen von balancierender Selektion gekoppelt sind, werden ebenfalls vermehrt Polymorphismen akkumuliert (Hudson and Kaplan, 1988 ; Hasselmann et al., 2008 (1)). Die Rekombination innerhalb des SDL ist stark unterdrückt (Hasselmann and Beye, 2006). Innerhalb des *csd*-Gens wurden trotzdem sporadische Rekombinationsereignisse in der Exon 2-5 Region detektiert (Hasselmann and Beye, 2006). Die Exone 6 bis 9 zeigen keine Anzeichen von Rekombination. Da Rekombination homogenisierend wirkt, würde durch sie allelische Spezifität zerstört werden (Hasselmann and Beye, 2006). Durch das Fehlen von Rekombinationsereignissen in den Exonen 6 bis 9 werden allelische Aminosäureunterschiede und Spezifität bewahrt (Hasselmann and Beye, 2006), was die Annahme nahelegt, dass die spezifitätsgebende Region von *csd* innerhalb dieser Exone zu finden ist (Hasselmann and Beye, 2006). Ob es zusätzlich zu der identifizierten PSD noch weitere Regionen oder Einzelpositionen im *csd*-Gen gibt, die für allelspezifische Spezifität verantwortlich sind, ist

bisher ungeklärt. Ebenso unbekannt ist, welche Aminosäuren der PSD für die Spezifität von *csd*-Allelen verantwortlich sind.

I.5 Zielsetzung der Arbeit

In einem Kreuzungsexperiment wurden insgesamt 15 *csd*-Allele in 23 Kreuzungen miteinander kombiniert. Das Ziel ist es, herauszufinden, welche der experimentell kombinierten *csd*-Allele funktionell verschieden sind. Außerdem wird untersucht, ob homozygote Allele, die in befruchteten Eiern vorhanden sind, in den adulten Arbeiterinnen fehlen. Dies wird mittels Bestimmung der *csd*-Allele von Arbeiterinnen und befruchteten Eiern aus den Kreuzungen erreicht.

Die Nukleotidsequenzen der 15 *csd*-Allele und die daraus resultierenden Aminosäuresequenzen werden ermittelt. Die Aminosäureunterschiede der experimentell kombinierten *csd*-Allele werden unter Verwendung verschiedener Datenmatrices, die aus den Sequenzdaten generiert wurden, mit Hilfe von Zwei-Weg hierarchischer Clusteranalyse (ZWHC) und Multidimensionaler Skalierung (MDS) herausgearbeitet. Das Ziel dabei ist, Aminosäureunterschiede zu identifizieren, die potentiell für die Spezifität von *csd*-Allelen verantwortlich sind, und herauszufinden, ob evtl. Kombinationen von Aminosäurepositionen mit Spezifität assoziiert sind. Dabei wird auch untersucht, wie groß die Veränderungen physikochemischer Eigenschaften sind, die durch Aminosäuresubstitutionen hervorgerufen werden und ob die physikochemischen Veränderungen ausreichend dafür sind, die Spezifität von *csd*-Allelen potentiell beeinflussen zu können. Die hypervariable Region (HVR) wird für diese Analysen aus der Datenmatrix entfernt und separat ausgewertet. Die hypervariablen Regionen werden dabei in Bezug auf Zusammensetzung und Länge in verschiedene Typen unterteilt. Es wird festgestellt, ob und wie Zusammenhänge zwischen den verschiedenen HVR-Typen und den Allelen, die mit Hilfe der ZWHC und der MDS beschrieben wurden, bestehen.

Mit Hilfe der Analysen werden Anhaltspunkte für *in vivo* Versuche gewonnen, mit denen gezielt Aminosäurepositionen auf ihre Bedeutung für die Spezifität von *csd*-Allelen untersucht werden können.

II. Material und Methoden

II.1 Chemikalien und Kitsysteme

Die verwendeten Chemikalien wurden hauptsächlich von den Firmen Merck (Darmstadt), Roth (Karlsruhe) und Sigma-Aldrich (Steinheim) bezogen und entsprachen dem Reinheitsgrad ‚pro analysis‘. Enzyme für molekularbiologische Arbeiten stammten von den Firmen Fermentas (St. Leon-Rot), New England Biolabs (Frankfurt a. M.), Promega (Mannheim) und Peqlab (Erlangen).

Radioaktive Nukleotide wurden von der Firma Hartmann Analytics (Braunschweig) bezogen.

Alle für molekularbiologische Arbeiten verwendeten Lösungen und Puffer wurden mit doppelt destilliertem Wasser angesetzt und, soweit erforderlich, autoklaviert, mit DEPC behandelt oder steril filtriert.

Kitsysteme:

DynabeadsTM mRNA Purification Kit, Dynal Biotech ASA, Oslo (Norwegen)
RevertAidTM Minus First Strand cDNA Synthesis Kit, Fermentas, St. Leon-Rot
FirstChoiceTM RLM-RACE Kit, Ambion, Austin (USA)
pGEMTM -T bzw. pGEMTM -T easy Vektor System, Promega, Mannheim
GenEluteTM Plasmid Miniprep Kit, Sigma-Aldrich, Steinheim
Qiagen PCR Purification Kit, Qiagen, Hilden
Peqlab Gold Gel Extraction Kit, Peqlab, Erlangen
E.Z.N.A. Cycle Pure Kit, Omega-Biotek, Norcross

Vektoren: Für die Transformation von PCR-Produkten wurden pGEMTM -T bzw. pGEMTM -T easy Vektoren mit Ampicilin-Resistenz und 3' T-Überhang verwendet.

Bakterienstämme:

E. coli Top 10TM

E. coli DH5 α

II.2 DNA-Extraktion

II.2.1 DNA-Extraktion aus Drohnen und Arbeiterinnen (modifiziert nach Hunt & Page 1995)

Adulte F1-Drohnen wurden nach in Gefrierbeuteln bei -20°C gelagert. Puppen der F2-Drohnen und F2-Arbeiterinnen und F2-Eier wurden in 70%igem Ethanol einzeln in Eppendorfreaktionsgefäßen bei -20°C gelagert.

Für die DNA Extraktion wurden Thoraces verwendet, die mit Pinzetten vom Bienenkörper getrennt und einzeln in Eppendorfreaktionsgefäße gegeben wurden. Im Falle der in Ethanol gelagerten Proben wurden die Thoraces 20 min in destilliertem Wasser bei RT im Thermoblock bei 600 rpm gewaschen.

Je ein Thorax wurde in ein Eppendorfreaktionsgefäß gegeben, 200 µl CTAB Lysispuffer und 3,5 µl Proteinase K zugefügt, mit einem Minimörser zerkleinert und für eine Stunde im Thermoblock bei 60°C und 600 rpm inkubiert. Für die Isolation der DNA wurde 300µl Phenol/Chloroform (zu gleichen Teilen) zugegeben, 5-6 mal invertiert und für 10 Minuten bei 13000 rpm zentrifugiert. Die entstandene obere Phase wurde abgezogen, in ein frisches Eppendorfreaktionsgefäß pipettiert, 200 µl Chloroform hinzugefügt, 5-6 mal invertiert und für zwei Minuten bei 13000 rpm zentrifugiert. Für die DNA Fällung wurde die entstandene obere Phase (ca. 250 µl) in ein frisches Eppendorfreaktionsgefäß überführt, 1/10 Volumen (25 µl) Natriumacetat und 2,4 Volumen (600µl) 96% iger Ethanol zugegeben, invertiert und zwei Stunden bei -20°C inkubiert. Danach wurde für 25 Minuten bei 13000 rpm und RT zentrifugiert, der Überstand abgezogen und verworfen und das entstandene Pellet mit 500 µl 70%igen Ethanol 3 min bei 13000 rpm gewaschen. Nach Entfernen des Überstandes wurde das Pellet bei Raumtemperatur oder bei 40°C im Thermoblock getrocknet und anschließend in 30 µl destilliertem Wasser gelöst.

II.2.2 DNA-Extraktion aus Einzeleiern

Für die DNA Extraktion aus Einzeleiern wurde obiges Protokoll leicht modifiziert:

Die bei -20°C in 70% igem Ethanol gelagerten Eier wurden einzeln in Eppis überführt und 2-3 min in destilliertem Wasser gespült. Die Homogenisation fand in 50 µl CTAB und 3µl Proteinase K statt. 70 µl Phenol/Chloroform, 50 µl Chloroform, 5 µl Natriumacetat, 100 µl 96%iger Ethanol Pellet in 15 ml dest. Wasser gelöst.

CTAB-Lysispuffer:

1% CTAB (Cetyltrimethylammoniumbromid)

50 mM Tris-HCl, pH 8.0

10 mM EDTA

750 mM NaCl

Proteinase K: 20 mg/ml

Phenol/Chloroform (zu gleichen Teilen): frisch angesetzt oder bei -20°C gelagert

pH Phenol > 7.5

II.3 Total-RNA-Präparation aus Eierpools und cDNA-Synthese

RNA wurde aus Pools von 40 – 50 in 70%igem Ethanol bei -20°C gelagerten Eiern gewonnen.

Nach Entfernen des Ethanol kurzes Spülen der Eier mit destilliertem Wasser und Einfrieren in flüssigem Stickstoff

- je Eierpool 50 µl Trizol zugeben, 1 min vortexen, 5 min bei RT inkubieren, 1 min vortexen
- 10 µl Chloroform zugeben, 30 sec vortexen, 5 min bei RT inkubieren und 15 min zentrifugieren
- Überstand abnehmen und in ein neues Eppendorfreaktionsgefäß überführen
- das gleiche Volumen (ca. 20 µl je Eierpool) Isopropanol zugeben, 4-6x invertieren und 10 min bei RT inkubieren

- 10 min bei 4°C zentrifugieren, Überstand verwerfen und RNA-Pellet mit 50 µl 70% Ethanol waschen
- 10 min bei 4°C zentrifugieren, Pellet trocknen und in 10 µl destilliertem Wasser resuspendieren

Lösungen:

Trizol-Extraktionspuffer:

0.8 M Guanidin-Thiocyanat

0.4 M Ammonium-Thiocyanat

0.1 M Natriumacetat, pH 5.0

5% Glycerin

38% Phenol

Die Isolation der mRNA aus den Total-RNA-Präparationen erfolgte mit Dynabeads nach dem Protokoll des Herstellers.

Die cDNA-Synthese aus mRNA erfolgte in zwei Schritten:

1. Erststrangsynthese:

- 11µl mRNA
- 1 µl Oligo dT Primer (oder random Hexamer Primer)
- vorsichtig mischen, in Microzentrifuge runter zentrifugieren
- 5 min 70°C, auf Eis abkühlen
- Tröpfchen durch Zentrifugation vereinigen, auf Eis
- 4 µl 5x Reaction Buffer
- 1 µl Ribonuklease Inhibitor(40 U/µl)
- 2 µl 10 mM dNTP- Mischung, Tröpfchen zentrifugieren
- bei Random Primern 5 min 25°C
- bei Oligo dT Primern 5 min 37°C
- 1µl Reversed Aid H (Minus) m-MuLV reverse Transkriptase (200 U/µl)
- bei Random Primern 10 min 25°C, 60 min 42°C

- bei Oligo dT Primern 60 min 42°C
- Hitzinaktivierung (10 min 70°C), abkühlen auf Eis

2. Zweitstrangsynthese:

- 20 µl Erststrang cDNA
- 10 µl 10x DANN-Polymerase I Buffer
- 4 µl DNA-Polymerase I
- 0,32 µl RNase H
- ad 100 µl (mit nukleasefreiem Wasser)
- 2 h 16°C

Aufreinigung der doppelsträngigen cDNA mit dem Gel-Elutions Kit von Eppendorf:

- 100 µl cDNA
- 500 µl Optinal Washing Buffer
- 100 µl Isopropanol
- mischen, über die Säule geben und nach Herstellerangaben zentrifugiert
- zweimal Elution mit 30 µl Elutionspuffer

II.4 PCR-Reaktionen und Restriction-Fragment-Length-Polymorphism (RFLP)-Analyse

Basis- PCR-Ansatz (10µl)

- 1 µl PCR-Puffer
- 1 µl dNTP-Mix (2mM)
- 0,5 µl reversed Primer (10µM)
- 0,5 µl forward Primer (10µM)
- 0,1 µl Taq-Polymerase
- 5,9 µl H₂O
- 1 µl DNA

Die DNA wurde 1:10 bis 1:50 verdünnt (entsprechend 10-50 ng/µl).

PCR Puffer:

- Tris 100 mM
- KCl 500 mM
- MgCl₂ 20 mM

Basis-PCR-Programm

94°C → 2min 30 sec (Anfangsdenaturierung)

94°C → 30 sec (zyklische Denaturierung)

50 °C → 40 sec (Annealing)

72 °C → 45 sec (zyklische Elongation)

35-37 Zyklen

72 °C → 2 min (End-Elongation)

Die Annealingtemperatur variierte je nach verwendetem Primerpaar (siehe Tabelle A.I bis A.IV). Wurde ein Primerpaar mit unterschiedlichen Annealingtemperaturen benutzt, wurde die niedrigere Temperatur gewählt. Die Elongationsdauer variierte je nach Länge des zu amplifizierenden Fragmentes.

Klonierungs-PCR

Für Klonierungen wurden 50 µl Ansätze des Basis-PCR -Ansatzes verwendet und in manchen Fällen die Taq durch ein Gemisch von Pwo:Taq (1:10 oder 1:20) ersetzt.

Für Exon 6-9 Fragmente betrug die Elongationszeit 45 sec, für Klonierungen des gesamten ORFs 1 min 20 sec. Die End-Elongation bei Klonierungs-PCRs betrug 15 min.

RFLP-Analyse

15 µl PCR-Ansatz

+10 µl Restriktionsmix (konventionelle Restriktionsendonukleasen):

- 2 µl Restriktionspuffer
- 0,2 µl Restriktionsendonuklease (10u/µl) (oder je 0,15 µl Restriktionsendonuklease 1 und Restriktionsendonuklease 2 bei Doppelrestriktionen)
- 7,8 µl H₂O

6 Stunden Inkubation bei 37°C

verwendete Restriktionsendonukleasen: Ssp I, Apo I und Mbo II

Die Restriktionen wurden auf ein Synergel oder ein 2 %iges Agarosegel aufgetragen. Die Gele liefen bei 75-90 V 3-4 Stunden, wurden in einem Ethidiumbromidbad gefärbt und auf einem UV-Tisch mit einer Kamera fotografiert.

Radioaktive PCR

Volumen für einen Ansatz (8 µl):

- 0,8 µl PCR-Puffer
- 0,8 µl dNTP (dCTP, dGTP, dTTP je 2mM, dATP 400 µM)
- 0,3 µl fw primer (10µM)
- 0,3 µl rev primer (10µM)
- 0,2 µl Taq Polymerase
- 0,2 µl [α -P³³] dATP (10 mCi/ml)
- 1 µl DNA
- 4,4 µl H₂O

Die DNA wurde 1:10 oder 1:20 verdünnt. Primer sind in Tabelle A.V angegeben.

Programm für radioactive PCR

94°C → 3min	(Anfangsdenaturierung)	
94°C → 35 sec	(zyklische Denaturierung)	37 Zyklen
50 °C → 1 min	(Annealing)	
72 °C → 1 min	(zyklische Elongation)	
72 °C → 2 min	(End-Elongation)	

Marker:

- Markerlösung:
- 3 µl DNA-Leiter (500 ng/ µl)
- 4,5 µl H₂O
-

radioactive Markierung der Markerlösung

- 1 µl [γ -P³³] ATP (10 mCi/ml)
- 1 µl 10 x T4-Kinase Puffer A
- 1 µl T4-Kinase
- 1,5 µl Markerlösung
- 5,5 µl H₂O
- → 37°C (120-150 min)
- 10 µl Formamide-Dye zugeben

Formamid-Loadingdye (2x):

- Volumen für 1 ml:
- 980 µl Formamid
- 20 µl 0,5 M EDTA
- 0,5 µg Xylene cyanol FF
- 0,5 µg Bromphenolblau

RFLP-Analyse des radioaktiven PCR-Produktes

Restriktion

- Reaktion mit einer Restriktionsendonuklease (Fast Digest):
- 0,3 µl Restriktionsendonuklease
- 1,4 µl Fast-Digest Puffer
- 5,3 µl H₂O
-
- Reaktion mit zwei Restriktionsendonukleasen (Fast Digest):
- 0,25 µl Restriktionsendonuklease 1
- 0,25 µl Restriktionsendonuklease 2
- 1,4 µl Fast-digest Puffer
- 5,1 µl H₂O
-
- 30 min 37 °C
- Hitzeinaktivierung (8 min 80°C)

Gelelektrophorese

- Proben mit 13 µl Formamid-Loadingdye versetzen
- Marker und Proben 2 min bei 94 °C denaturieren
- auf Eis stellen und sofort auf ein denaturierendes Harnstoffgel auftragen (6,2 µl Proben, 5,5 µl Marker)
- 1350 V, 170-180 min
- Übertragen des Gels auf ein Filterpapier und Trocknung bei 80°C, 90-120 min in einem Gelrockner
- Exposition eines Röntgenfilms für 20-24 h (-70°C)
- Entwicklung des Films im Fotolabor

denaturierendes Harnstoffgel

- 66 ml Harnstofflösung (54,5 %)
- 10 ml Acrylamid (19:1)
- 4 ml 10xTBE
- 560 µl APS (10%)
- 32 µl TEMED

Lösungen

- 10xTBE: 108 g Trisbase, 55g Borsäure, 40 ml 0,5 M EDTA, H₂O ad 1l, pH 8
- APS: 1g Ammoniumpersulfat, H₂O ad 10 ml

Glasplatten:

kleine Platte: ca. 39 x 33 cm Ausmaße

große Platte: ca. 41 x 33 cm

Spacer: 0,35 mm

Ein Gel hat die Maße von ca. 39 x 31 x 0,0035 cm.

Die Glasplatten wurden mit Dimethyl- Dichlorsilan beschichtet. Nach dem Gießen polymerisierte das Gel für zwei Stunden, bevor es beladen oder über Nacht im Kühlschrank gelagert wurde. Vor dem Beladen musste das Gel 45-60 min bei 1350 V vorlaufen, bis es sich gleichmäßig auf 60°C erwärmt hatte.

II.5 Klonierungen

Die Sequenzen der Exon 6-9 Regionen (F1- und F2-Drohnen) und die 5'UTR-Exon 6/7/8 Regionen (F2-Eiern) wurden durch Klonierungen und anschließende Sequenzierungen der Klone ermittelt.

Gewinnung der Inserts

- Auftrennung der PCR-Produkte mittels Gelektrophorese (1%iges Agarosegel, 120 V, 60-70 min)
- Färbung im Ethidium-Bromid Bad
- sauberes Ausschneiden des gewünschten Fragmentes aus dem Gel
- Elution des PCR-Produktes aus dem Gel mit dem Gelelutions-Kit
- Mengenbestimmung der Inserts durch Vergleich mit 5 ng mit HindIII geschnittener λ -DNA auf einem Agarosegel

Ligation

- 25 ng pGEMTM-T Vektor
- 4-fache Menge Insert
- Ansatz der Ligation laut Herstellerangaben, bezogen auf 25 ng Vektor
- Inkubation bei 17°C ü.N.

Transformation

- kompetente E. Coli (50µl Aliquots) auf Eis auftauen
- Ligationsansatz zugeben, 20 min auf Eis
- 50 sec Hitzeschock (42°C)
- 2 min auf Eis
- Zugabe von 450 µl LB-Medium, Inkubation für 90-120 min bei 37°C und 180 rpm auf dem Schüttler
- Ausplattieren auf LB-Amp Platten und Inkubation bei 37°C, ü.N.

1. Klonierungen der Exon 6-9 Sequenzen von F1- und F2-Drohnen

Es wurden von den Exon 6-9 Sequenzen der F1-bzw. F2-Drohnen der verschiedenen Kreuzungen je 2-3 Klone erstellt und zusätzlich PCR-Produkte sequenziert, um die Richtigkeit der Klone zu überprüfen. Für Sequenz und Lage der verwendeten Primer siehe Tabelle A.I im Anhang.

2. Klonierungen der 5' UTR bis Exon 6/7/8 Sequenzen

Für die Klonierung der 5' UTR bis Exon 6/7/8 Regionen wurden 40-50 F2-Eier jeweils einer Inzuchtkreuzung vereinigt und deren mRNA mittels reverser Transkriptase in cDNA umgeschrieben. In jedem Pool befanden sich drei *csd*-Allele. Um gezielt die einzelnen Allele amplifizieren zu können, wurden mit Hilfe der Exon 6-9 Sequenzen reversed Primer generiert, die spezifisch jeweils nur eines der Allele in den Exonen 6 oder 7 oder 8 binden konnten (für Lage und Sequenz siehe Tabelle A.III im Anhang). Als forward Primer wurden konservierte, im 5'-Bereich liegende Oligonukleotide (Tabelle A.II im Anhang) benutzt oder eine RACE durchgeführt Primersequenzen in Tabelle A.IV im Anhang).

Von jedem Allel wurden mindestens drei Klone hergestellt, die durch Vergleich mit den Exon6-9 Sequenzen der jeweiligen Allele auf Richtigkeit überprüft wurden.

II.6 Experimenteller Aufbau

II.6.1. Kreuzungsexperiment

Das Kreuzungsexperiment wurde von M. K. Fondrk und R.E. Page durchgeführt. In der Parentalgeneration wurden 5 Königinnen (B, G, R, W und Y) einzelbesamt mit Spermien von 5 verschiedenen Drohnen. Diese 5 Kreuzungen wurden nach der P-Königin benannt und als B, G, R, W und Y-Linie bezeichnet. Da mit jeder P-Kreuzung drei Allele eingebracht wurden, befanden sich im gesamten Kreuzungsexperiment 15 *csd*-Sequenzen (siehe Abbildung I.2.1). Die Spermien der F1-Drohnen jeder Kreuzung wurden gesammelt und vereinigt. Die fünf entstandenen Spermienpools enthielten jeweils die beiden Allele der P-Königin und wurden nach der entsprechenden Linie benannt (Spermienpool B, G, R, W und Y). Die Königinnen der F1-Generation können je Linie in zwei verschiedenen Genotypen auftreten. Sie tragen eines der beiden Allele der P-Königin und das P-Drohnen-Allel. In den F1-Kreuzungen wurden 24 F1-Königinnen mit je einem der fünf Spermienpools inseminiert (Tabelle I.2.1). Die F1-Drohnen, F2-Drohnen, F2-Arbeiterinnen und F2-Eier wurden gesammelt und in -20°C gelagert.

Tabelle II.1: Inseminierungen von F1-Königinnen mit Spermienpools der F1 Drohnen.

Kreuzung	F1-Königin	F1-Spermienmix
B05	B	G
B10	B	R
B12	B	B
G07	G	G
G08	G	G
G10	G	Y
G12	G	R
G14	G	B
R06	R	G
R11	R	R
R13	R	B
R14	R	W
R8	R	Y
W40	W	G
W41	W	Y
W43	W	R
W45	W	B
W47	W	W
Y06	Y	G
Y10	Y	Y
Y12	Y	R
Y13	Y	B
Y15	Y	W

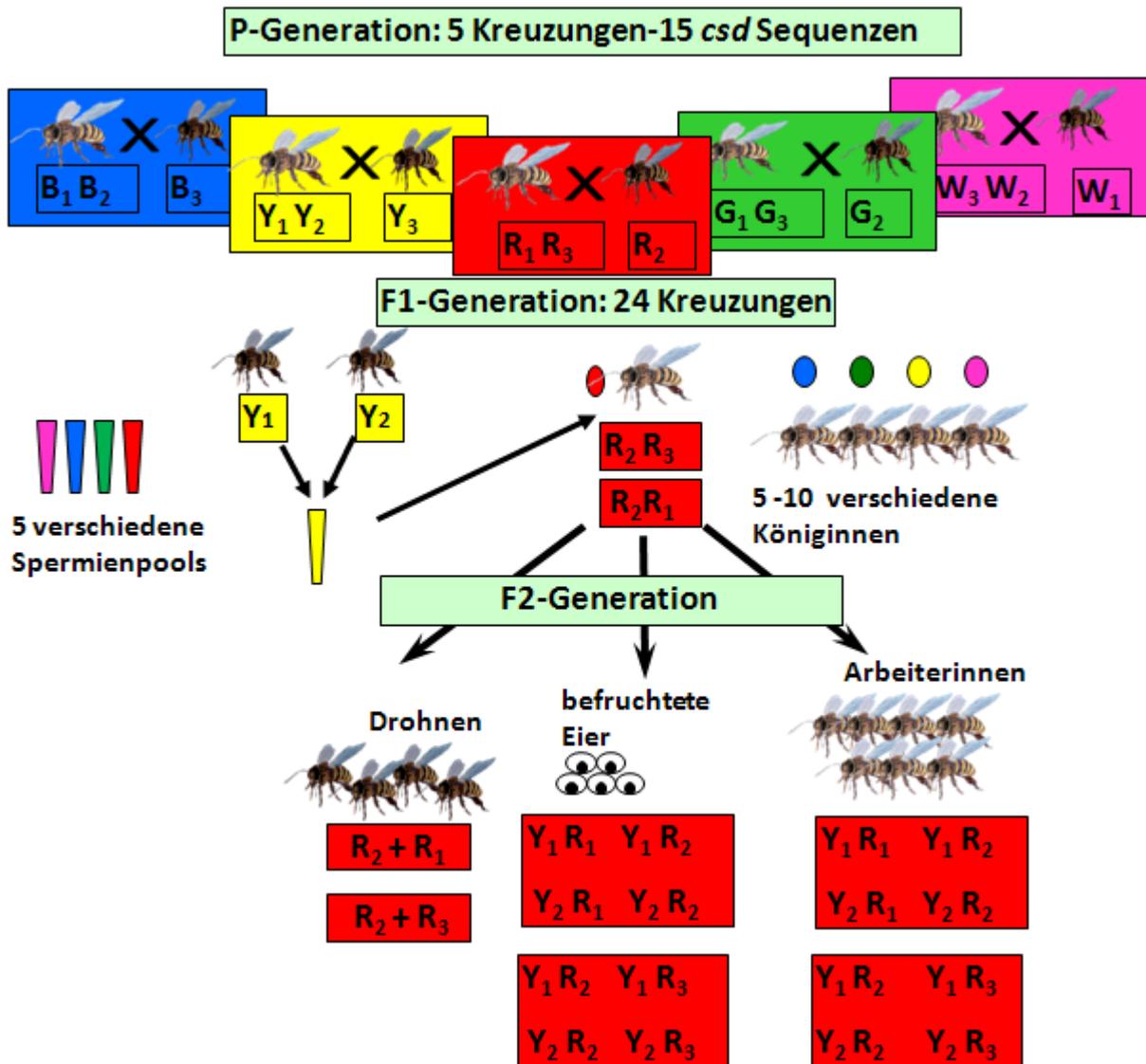


Abb. Abbildung II.1: Kreuzungsschema der P- und der F1-Kreuzungen

Die P-Königinnen der verschiedenen Linien (B, G, R, W und Y) werden mit je einem P-Drohn besamt. Die Drohnen der F1-Generation tragen je eines der beiden Allele der P-Königin. Diese beiden Allele befinden sich in den F1-Spermienpools der jeweiligen Linien. Für die F1-Königinnen gibt es je Linie zwei mögliche Genotypen. Sie tragen je eines der beiden Allele der P-Königin und das Allel des P-Drohns. Die F2-Drohnen tragen je eines der beiden Allele der F1-Königin. In den befruchteten F2-Eiern existieren vier *csd*-Kombinationen, unabhängig davon, ob sich die kombinierten Sequenzen funktionell unterscheiden. In den F2-Arbeiterinnen treten nur diejenigen Kombinationen auf, die aus funktionell verschiedenen *csd*-Allelen bestehen. Dargestellt ist die F1-Kreuzung R8.

Unter der Annahme, dass die Wahrscheinlichkeit dafür, eine Allelkombination zu finden, für alle vier Kombinationen gleich ist ($P = 0,25$), lässt sich mit Hilfe der Multinomialverteilung die Anzahl der Arbeiterinnen bestimmen, bei der mit 95%iger Wahrscheinlichkeit jede Kombination mindestens einmal vorkommt.

$$P(M) = \frac{4^n - 4 \cdot 3^n + 6 \cdot 2^n - 4}{4^n}$$

M: Ereignis, dass jede Kombination mindestens einmal vorkommt

$P(M)$: Wahrscheinlichkeit, dass das Ereignis M eintritt

Einsetzen von Werten für n zeigt nun:

n = 15: $P(M) \approx 0,94673 < 0,95$,

n = 16: $P(M) \approx 0,96 > 0,95$.

Werden also 16 oder mehr Arbeiterinnen pro Kreuzung getestet, so wird mit einer Wahrscheinlichkeit von mehr als 95% von jeder Kombination mindestens eine gefunden.

Werden mehr als 16 Arbeiterinnen untersucht und nicht alle Kombinationen gefunden, so kann davon ausgegangen werden, dass die nichtgefundenen Kombinationen nicht existieren und die entsprechenden Allele funktionell gleich sind. Die Tests wurden mittels radioaktiver PCR und anschließender RFLP- Analyse durchgeführt.

II.6.2 Bestimmung der Consensussequenzen der klonierten Nukleotidsequenzen mit Hilfe des Programmes GAP (Staden Package)

Die Sequenzdaten der Exon 6-9 - Klone und die 5'-UTR-Exon 6/7/8 cDNA-Klone wurden in GAP in einem Alignment angeordnet (normal shotgun assembly) und die Consensussequenzen der jeweiligen Allele ermittelt.

II.6.3 Zusammensetzen der *csd*-Sequenzen aus den Klonierungen der 5' UTR bis Exon 6/7/8 Regionen und der Exon 6-9 Region

Die Consensussequenzen der Exon 6-9 Sequenzen wurden in Bioedit geöffnet, mittels der ClustalW Methode in einem Alignment angeordnet und die Introne manuell entfernt. Um aus den 5'-UTR-Exon 6/7/8- und den Exon 6-9 Sequenzen die Nukleotidsequenzen der Allele in voller Länge zusammensetzen zu können, wurden die beiden Consensussequenzen eines Allels in Bioedit geöffnet und der überlappende Bereich manuell herausgeschnitten. Das Alignment der vollständigen Open Reading Frames enthielt 431 Aminosäurepositionen bzw. 1293 Nukleotidpositionen. Das vollständige Aminosäurealignment ist in Abbildung A.I im Anhang angegeben.

II.6.4 Untersuchung der phylogenetischen Verwandtschaft der *csd*-Allele mit Hilfe des Programmes Molecular Evolutionary Genetics Analysis (MEGA)

Für die Analyse wurde ein Alignment der Nukleotidsequenzen der 14 *csd*-Allele in voller Länge und ein Alignment der genomischen Exon 6-9 Region genutzt. Die hypervariable Region (Aminosäurepositionen 325 bis 351 des Alignments und die entsprechenden Nukleotide in dem Exon 6-9 Alignment) sowie alle Deletionen in den Alignments wurden entfernt.

Mit Hilfe des Programmes DnaSp 5.0 wurden die variablen Nukleotidpositionen der Introne in der genomischen Exon 6-9 Region und die synonymen Nukleotidsubstitutionen in den Nukleotidsequenzen der *csd*-Allele in voller Länge extrahiert. Die variablen Nukleotidpositionen der Introne und die synonymen Nukleotidsubstitutionen in den Nukleotidsequenzen der *csd*-Allele in voller Länge wurden in MS Office Word vereinigt. Aus diesen, aus polymorphen Nukleotidpositionen bestehenden Nukleotidsequenzen wurde mittels des Programmes MEGA ein Ähnlichkeitsbaum rekonstruiert, mit dessen Hilfe die phylogenetische Verwandtschaft der Allele analysiert wurde. Der Ähnlichkeitsbaum wurde mit Hilfe der Neighbour-Joining Methode erstellt. Der Test auf Phylogenie erfolgte anhand der Bootstrap Methode (500 Wiederholungen).

II.7 Erstellung von Datenmatrices für die Analyse mit Zwei-Weg hierarchischer Clusteranalyse und Multidimensionaler Skalierung

II.7.1. Erstellen einer Datenmatrix basierend auf Aminosäureunterschieden: Matrix 1

Um die Unterschiede in den Aminosäuresequenzen der 14 *csd*-Allele den Analysen mit Zwei-Weg hierarchischer Clusteranalyse und Multidimensionale Skalierung zugänglich zu machen, werden die Informationen über die Aminosäureunterschiede der 71 paarweisen Sequenzkombinationen in Datenmatrices transformiert. Hierfür wird ein Alignment der Aminosäuresequenzen der 14 *csd*-Allele in Bioedit erstellt. Aus dem 431 Aminosäurepositionen enthaltenden Alignment wird die hypervariable Region (Aminosäurepositionen 325 bis 351) entfernt, so dass 404 Positionen verbleiben. Es folgen paarweise Vergleiche der 71 Sequenzkombinationen. Jede Position bekommt in jedem paarweisen Vergleich einen Wert. Wenn sich die Aminosäuren der beiden Sequenzen an der betrachteten Position unterscheiden beträgt der Wert 1, wenn sie sich nicht unterscheiden 0 (siehe Tabelle II.2.A). Als Unterschied wird auch gewertet, wenn bei einer der beiden Sequenzen eine Aminosäure steht und die andere Sequenz an dieser Position eine Deletion aufweist.

Um die für jedes Allel typischen Positionsmuster zu ermitteln, werden die paarweisen Vergleiche, in denen das jeweilige Allel vorkommt, untereinander geschrieben und der Mittelwert jeder Position errechnet (siehe Tabelle II.2.B). Jeder Vergleich wird daher bei der Erstellung der Matrix insgesamt zweimal verwendet. Die Positionsmuster der 14 Allele werden untereinander geschrieben und Positionen, die bei allen Allelen gleich sind (Nullpositionen), gelöscht (siehe Tabelle II.3). Es verbleiben 82 Positionen, die Aminosäuredifferenzen zwischen den Allelen aufweisen.

Die Informationen, die in den 71 Vergleichen stecken, werden so auf die Positionsmuster der 14 *csd*-Allele übertragen. Die entstandene Datenmatrix enthält 14 Reihen (*csd*-Allele) und 82 Spalten (Positionen). Diese 14x82 - Matrix wird im Folgenden als Matrix 1 (M1) bezeichnet. Durch Vergleich der Reihen kann die Ähnlichkeit der Allele zueinander analysiert werden und durch Vergleich der Spalten die der Positionen. Tabelle II.2 zeigt die Berechnung der

Matrixwerte am Beispiel des Allels B3. Tabelle II.3 zeigt die ersten acht Aminosäurepositionen der Matrix 1.

Tabelle II.2. A und B: Transformation von Sequenzdaten in eine Datenmatrix am Beispiel der Sequenz B3

A: Vergleich von Position 1 bis 9 der Aminosäuresequenzen von B1 und B3. Positionen, die keine Aminosäuredifferenzen aufweisen, bekommen den Wert 0. Positionen, die Aminosäuredifferenzen aufweisen bekommen den Wert 1.

	Positionen								
	1	2	3	4	5	6	7	8	9
B1	M	K	R	N	I	S	S	Y	S
B3	M	K	R	N	I	S	N	Y	S
Sequenzvergleich B3B1	0	0	0	0	0	0	1	0	0

B: Berechnung der Positionswerte von B3. Alle Sequenzvergleiche in denen B3 vorkommt werden untereinander geschrieben und die für das Allel B3 typischen Positionswerte durch Mittelwertbildung berechnet. Dargestellt sind die ersten 9 Positionswerte der Vergleiche, bei denen B3 beteiligt ist und das für das Allel B3 typische Positionsmusters.

		Positionen								
		1	2	3	4	5	6	7	8	9
B3-Sequenzvergleiche	B3B1	0	0	0	0	0	0	1	0	0
	B3B2	0	0	0	0	0	0	0	0	0
	B3G1	0	0	0	0	0	0	1	0	0
	B3G3	0	0	0	0	0	0	1	0	0
	B3R1	0	0	0	0	0	0	0	0	0
	B3R3	0	0	0	0	0	0	0	0	0
	B3-Positionsmuster		0	0	0	0	0	0	0,5	0

Tabelle II.3: Auszug aus Datenmatrix M1. Dargestellt sind die ersten 8 Positionswerte der Positionsmuster der 14 Allele nach Entfernen der Nullpositionen.

		Position							
		7	23	24	42	54	57	83	93
Sequenz	B1	0,727	0,727	0,727	0,636	0,727	0,091	0,091	0,182
	B2	0,455	0,455	0,455	0,545	0,455	1,000	0,091	0,273
	B3	0,500	0,500	0,500	0,667	0,500	0,167	0,000	0,333
	G1	0,667	0,667	0,667	0,667	0,667	0,111	0,111	0,778
	G2	0,500	0,500	0,500	0,583	0,500	0,083	0,083	0,250
	G3	0,583	0,583	0,583	0,500	0,583	0,083	0,083	0,833
	R1	0,545	0,545	0,545	0,636	0,545	0,000	0,091	0,273
	R2	0,600	0,600	0,600	0,500	0,600	0,100	1,000	0,800
	R3	0,455	0,455	0,455	0,545	0,455	0,091	0,091	0,182
	W1	0,600	0,600	0,600	0,500	0,600	0,100	0,000	0,200
	W2	0,375	0,375	0,375	0,500	0,375	0,125	0,125	0,125
	Y1	0,667	0,667	0,667	0,444	0,667	0,000	0,111	0,222
	Y2	0,545	0,545	0,545	0,455	0,545	0,091	0,091	0,273
		Y3	0,400	0,400	0,400	0,500	0,400	0,100	0,000

Die komplette Matrix 1 befindet sich in Tabelle A.VI im Anhang.

II.7.2. Erstellen von Datenmatrices basierend auf physikochemischen Differenzen von substituierten Aminosäuren: Matrix 2 und Matrix 3

Eine weitere Möglichkeit, die Verschiedenheit zweier Allele darzustellen ist, die Unterschiede in den physikochemischen Eigenschaften, die durch Aminosäuresubstitutionen in einer Sequenz entstehen, in einer Datenmatrix zu quantifizieren. Die physikochemischen Differenzen zwischen den 190 möglichen Aminosäurepaaren werden nach Grantham (1974) bestimmt (Tabelle II.5). In die Berechnung gehen Zusammensetzung (c) der Aminosäure, Polarität (p) und Volumen (v) ein (Tabelle II.4). Der Wert für die Zusammensetzung wird berechnet aus dem Verhältnis der Atomgewichte von Hetero(Nichtkohlenstoff)elementen in Endgruppen oder Ringen zu Kohlenstoffatomen in der Seitenkette. Die Werte für Polarität und molekulares Volumen stammen aus publizierten Daten (C.R. Woese, (1973), A. A. Aboderin (1971), D. E. Goldsack and R. C. Chalifoux, J. (1973)). Die physikochemische Differenz ist eine einheitsfreie Kenngröße.

Tabelle II.4: physikochemische Eigenschaften der Aminosäuren.

Aminosäure	Abk.	Eigenschaft		
		Zusammensetzung	Polarität	mol. Volumen (v)
Serin	S	1,42	9,2	32
Arginin	R	0,65	10,5	124
Leucin	L	0	4,9	111
Prolin	P	0,39	8	32,5
Threonin	T	0,71	8,6	61
Alanin	A	0	8,1	31
Valin	V	0	5,9	84
Glycin	G	0,74	9	3
Isoleucin	I	0	5,2	111
Phenylalanin	F	0	5,2	132
Tyrosin	Y	0,2	6,2	136
Cystein	C	2,75	5,5	55
Histidin	H	0,58	10,4	96
Glutamin	Q	0,89	10,5	85
Asparagin	N	1,33	11,6	56
Lysin	K	0,33	11,3	119
Asparaginsäure	D	1,38	13	54
Glutaminsäure	E	0,92	12,3	83
Methionin	M	0	5,7	105
Tryptophan	W	0,13	5,4	170
Durchschnittliche chemische / physikalische Distanz \bar{D}		0,739	3,134	50,06

Die durchschnittliche chemische / physikalische Distanz für die einzelnen Eigenschaften (c oder p oder v) wurde mit folgender Formel berechnet:

$$\bar{D}_x = \sqrt{\sum[(x_i - x_j)^2] / 190} \text{ mit } x = c \text{ oder } p \text{ oder } v$$

Die physikochemischen Differenzen der einzelnen Aminosäurepaare wurden mit folgender Formel berechnet:

$$D_{ij} = \left[\alpha(c_i - c_j)^2 + \beta(p_i - p_j)^2 + \gamma(v_i - v_j)^2 \right]^{\frac{1}{2}} \times 50,723$$

Tabelle II.5: Physikochemische Differenz D für jedes Aminosäurepaar

Abk.	R	L	P	T	A	V	G	I	F	Y	C	H	Q	N	K	D	E	M	W
S	110	145	74	58	99	124	56	142	155	144	112	89	68	46	121	65	80	135	177
R		102	103	71	112	96	125	97	97	77	180	29	43	86	26	96	54	91	101
L			98	92	96	32	138	5	22	36	198	99	113	153	107	172	138	15	61
P				38	27	68	42	95	114	110	169	77	76	91	103	108	93	87	147
T					58	69	59	89	103	92	149	47	42	65	78	85	65	81	128
A						64	60	94	113	112	195	86	91	111	106	126	107	84	148
V							109	29	50	55	192	84	96	133	97	152	121	21	88
G								135	153	147	159	98	87	80	127	94	98	127	184
I									21	33	198	94	109	149	102	168	134	10	61
F										22	205	100	116	158	102	177	140	28	40
Y											194	83	99	143	85	160	12	36	37
C												174	154	139	202	154	170	196	215
H													24	68	32	81	40	87	115
Q														46	53	51	29	101	130
N															94	23	42	142	174
K																101	56	95	110
D																	45	160	181
E																		126	152
M																			67

Wie bei Erstellung von Matrix 1 folgen paarweise Vergleiche der 71 Sequenzkombinationen. Jede Position bekommt in jedem paarweisen Vergleich einen Wert für die physikochemische Differenz der beiden Aminosäuren an dieser Position. An den Positionen 144, 211, 307, 358, 359, 362, 372, 373, 419, 431 treten Aminosäuredeletionen auf. Aminosäurepositionen, die bei bestimmten Allelen Deletionen aufweisen, werden bei den paarweisen Vergleichen der entsprechenden Allele nicht einbezogen und gehen deshalb nicht in Berechnung der Positionswerte der entsprechenden Allele ein. Bei den Positionen 358 und 417 kommt gar kein paarweiser Vergleich zustande, denn sie haben jeweils nur bei einem Allel eine Aminosäure. Diese beiden Positionen sind daher in der Matrix zwar vorhanden aber ohne Angaben von Positionswerten. Die anderen acht Aminosäurepositionen weisen nur bei einigen Allelen Deletionen auf und haben daher nur bei den entsprechenden Allelen keine Positionswerte. Die charakteristischen Positionswerte für jedes Allel werden dann wie bei M1 ermittelt und untereinander geschrieben. Die so entstandene Datenmatrix enthält 14 Reihen (*csd*-Allele) und 82 Spalten (Positionen) und wird als Matrix 2 (M2) bezeichnet. In Tabelle II.6 sind die ersten acht Positionswerte der 14 Sequenzen dargestellt. Die komplette Matrix 2 befindet sich in Tabelle A.VII im Anhang.

Tabelle II.6: Auszug aus Datenmatrix M2. Dargestellt sind die ersten 8 Positionswerte der 14 Sequenzen nach Entfernen der Nullpositionen, basierend auf den physikochemischen Differenzen der Aminosäuredifferenzen.

		Position							
		7	23	24	42	54	57	83	93
Sequenz	B1	33,64	33,64	41,63	41,00	21,29	0,92	6,45	10,64
	B2	21,02	21,02	26,02	35,14	13,31	10,12	6,45	15,97
	B3	23,13	23,13	28,62	42,95	14,64	1,69	0,00	19,52
	G1	30,83	30,83	38,16	42,95	19,52	1,12	7,89	45,54
	G2	23,13	23,13	28,62	37,58	14,64	0,84	5,91	14,64
	G3	26,98	26,98	33,39	32,22	17,08	0,84	5,91	48,79
	R1	25,23	25,23	31,22	41,00	15,97	0,00	6,45	15,97
	R2	27,75	27,75	34,34	32,22	17,57	1,01	70,97	46,84
	R3	21,02	21,02	26,02	35,14	13,31	0,92	6,45	10,64
	W1	27,75	27,75	34,34	32,22	17,57	1,01	0,00	11,71
	W2	17,34	17,34	21,46	32,22	10,98	1,27	8,87	7,32
	Y1	30,83	30,83	38,16	35,79	19,52	0,00	7,89	13,01
	Y2	25,23	25,23	31,22	35,14	15,97	0,92	6,45	15,97
	Y3	18,50	18,50	22,89	32,22	11,71	1,01	0,00	11,71

Die Matrices werden im Folgenden mit zwei verschiedenen Algorithmen (Zwei-Weg hierarchische Clusteranalyse und Multidimensionale Skalierung, siehe Abschnitt II.8) analysiert. Matrix 2 weist aufgrund der Aminosäuredeletionen Lücken auf. Ein lückenhafter Datensatz wird von dem Programm, mit dessen Hilfe die Multidimensionale Skalierung durchgeführt wird, nicht akzeptiert, von dem Programm, mit dessen Hilfe die Zwei-Weg hierarchische Clusteranalyse durchgeführt wird, schon. Um die Informationen über die physikochemischen Unterschiede dennoch mittels Multidimensionaler Skalierung auswerten zu können, wird eine weitere Matrix erstellt, in der die Positionen, die Deletionen aufweisen fehlen. Die entsprechende Matrix besteht aus 14 Reihen (*csd*-Allele) und 72 Spalten (Positionen) und wird als Matrix 3 (M3) bezeichnet. Matrix 3 befindet sich in Tabelle A.VIII im Anhang.

II.8 Datenanalyse mit Hilfe von Zwei-Weg hierarchischer Clusteranalyse und Multidimensionaler Skalierung

II.8.1 Transformation der Datenmatrices in Matrices, die die paarweise Ähnlichkeit der Daten darstellen (Ähnlichkeitsmatrices)

Bevor die Datenmatrices M1, M2 und M3 mit Hilfe der oben angegebenen Algorithmen ausgewertet werden können, müssen aus ihnen Matrices erstellt werden, die die paarweise Ähnlichkeit der Daten angeben. Dies sind entweder die paarweisen Ähnlichkeiten der Allele oder die paarweisen Ähnlichkeiten der Aminosäurepositionen. Paarweise Ähnlichkeiten können z.B. durch Korrelation der Daten oder durch Bildung der Euklidischen Distanzen ermittelt werden.

I.8.1.1 Berechnung einer Ähnlichkeitsmatrix mittels Spearman'scher Rangkorrelation.

Die Qualität des linearen Zusammenhangs zweier Datenreihen kann durch Korrelation der Daten bestimmt werden. Am geläufigsten ist die Korrelation mit Hilfe des Pearson'schen Produkt-Moment –Korrelationskoeffizienten r . Dafür müssen die Daten allerdings einer Normalverteilung unterliegen. Ist dies nicht der Fall, so kann die Korrelation der Daten mit Hilfe des Spearman'schen Rangkorrelationskoeffizienten r_s , einer Abwandlung des Pearson'schen Korrelationskoeffizienten, abgeschätzt werden. Der Unterschied dieser Rangkorrelation zur Korrelation nach Pearson besteht darin, dass die Werte durch Ränge ersetzt werden, bevor der Korrelationskoeffizient berechnet wird. Die Daten müssen dazu nicht normalverteilt sein.

Der Spearman'sche Rangkorrelationskoeffizient r_s zweier Datenreihen x und y ist definiert als:

$$r_s = \frac{\sum_{i=1}^n (rgx_i - \overline{rgx})(rgy_i - \overline{rgy})}{\sqrt{\sum_{i=1}^n (rgx_i - \overline{rgx})^2 \sum_{i=1}^n (rgy_i - \overline{rgy})^2}}$$

Mit

rgx_i bzw. rgy_i = Rang des Wertes von x_i bzw. y_i

\overline{rgx} bzw. \overline{rgy} = Mittelwert der Ränge von x bzw. y

n = Gesamtzahl der Werte einer Datenreihe.

Der Rangkorrelationskoeffizient r_s nimmt Werte zwischen +1 und -1 an. Je höher der Betrag von r_s desto größer ist der lineare Zusammenhang der verglichenen Datenreihen. Ein Rangkorrelationskoeffizient von 1 bedeutet vollständige positive lineare Abhängigkeit der beiden Datenreihen (je höher der x-Wert desto höher der y-Wert). Ein Wert von -1 bedeutet, dass zwischen den Datenreihen eine vollständige negative lineare Abhängigkeit besteht (je höher der x-Wert desto niedriger der y-Wert oder umgekehrt). Ein Rangkorrelationskoeffizient von 0 bedeutet, dass kein linearer Zusammenhang zwischen den Datenreihen besteht.

Die Matrix, die die Korrelationskoeffizienten der 14 *csd*-Allele darstellt, besteht aus 14 x 14 Datenpunkten. Die Matrix, die die Korrelation der Aminosäurepositionen darstellt, besteht aus 82 x 82 (M1 und M2) bzw. 72 x 72 Datenpunkten (M3). Jeder Datenpunkt der Matrix entspricht dem Korrelationskoeffizienten zweier *csd*-Allele bzw. zweier Aminosäurepositionen.

II.8.1.2 Berechnung einer Ähnlichkeitsmatrix mittels Bildung der Euklidischen Distanzen zwischen den Datenreihen

Die Euklidische Distanz zwischen zwei Datenreihen (Allele oder Aminosäurepositionen) nimmt den Unterschied zwischen den verglichenen Werten direkt. Im Gegensatz zu korrelationsbasierten Methoden der Distanzmessung wird das Ausmaß der Unterschiede zwischen den Werten angerechnet. Die Varianz der Werte spielt keine Rolle.

Die Euklidische Distanz d zwischen zwei Datenreihen x und y ist definiert als:

$$d = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2.$$

Mit

x_i bzw. y_i = Wert der Datenreihe x bzw. y an der Stelle i

n = Gesamtzahl der Werte einer Datenreihe.

Je geringer die Euklidische Distanz zweier Datenreihen ist, desto ähnlicher sind sie, je größer sie ist, desto unähnlicher.

Die Matrix, die die Distanzen der *csd* Allele darstellt besteht aus 14 x 14 Datenpunkten, die Matrix, die die Distanzen der Aminosäurepositionen darstellt besteht aus 82 x 82 bzw. 72 x 72 Datenpunkten. Jeder Datenpunkt der Matrix entspricht der Euklidischen Distanz zweier *csd* Allele bzw. zweier Aminosäurepositionen.

Für die Analyse der 14 *csd* Allele mittels ZWHC und MDS werden sowohl Ähnlichkeitsmatrices, die auf Korrelation der Werte beruhen, als auch Ähnlichkeitsmatrices, die auf Bestimmung der Euklidischen Distanzen beruhen, verwendet. Die (Un)Ähnlichkeitsmatrices der Allele und der Positionen aus Matrix 1 und 3 sind im Anhang angegeben (Tabellen A.XV bis A.XXII).

II.8.2 Analyse der (Un)ähnlichkeitsmatrices mittels Zwei-Weg hierarchischer Clusteranalyse und Multidimensionaler Skalierung

II.8.2.1 Zwei-Weg hierarchische Clusteranalyse mit Hilfe des Programms „Cluster 3.0“

Mit Hilfe des Programmes „Cluster 3.0“ können die Distanzmatrices der Allele und der Positionen parallel (Zwei-Weg) nach deren Ähnlichkeiten hierarchisch gruppiert werden. Das Ergebnis der Clusteranalyse kann dann mit Hilfe des Programms „Java Treeview“ dargestellt werden. Die Positionswerte haben ihrer Höhe entsprechend bei den verschiedenen Allelen eine unterschiedlich intensive Farbe. Je höher der Wert, desto intensiver der Farbton. Die Farbintensität kann in Java Treeview in einem Intervall von 0 bis maximal 5 variiert werden. Für Darstellungen der Distanzmatrices von Matrix 1 ist dies ausreichend, da hier ein Positionswert maximal den Wert 1 annehmen kann. Bei Matrix 2 und 3 liegen die Maximalwerte mit 225,37 und 195,74 außerhalb der Darstellungsbreite. Die Positionswerte von Matrix 2 bzw. Matrix 3 werden deswegen vor der Clusteranalyse durch 45,1 bzw. 39,15 geteilt, so dass die Maximalwerte bei 5 liegen.

II.8.2.2 Multidimensionale Skalierung (MDS) mit Hilfe des Programms „Primer 6“

Bei der Darstellung von Distanzen zwischen Allelen oder Aminosäurepositionen mittels Multidimensionaler Skalierung werden die Datenpunkte, die die Allele oder Aminosäurepositionen darstellen, in einem 2- oder 3- dimensionalen Raum so angeordnet, dass deren Abstände möglichst entsprechend den Abständen wiedergegeben werden, die sich aus der (Un)ähnlichkeitsmatrix ergeben. Je enger zwei Datenpunkte die die Allele oder die Aminosäurepositionen zusammen liegen, desto ähnlicher sind die Allele bzw. die Aminosäurepositionen. Das Ergebnis der Analyse wird sowohl graphisch als auch tabellarisch (Koordinatensystem) angegeben. Wie gut die Ordination die Abstände der Datenpunkte aus den (Un)Ähnlichkeitsmatrices wiedergibt, kann mit Hilfe der Stresswerte, die bei jeder Ordination angegeben werden, beurteilt werden. Clark and Warwick (2001) geben Anhaltspunkte für die Bewertung der Stresswerte: Stresswerte unter 0,1 bedeuten demnach, dass die Ordination gut ist. Bei Stresswerten unter 0,2 ist ein weniger exaktes Abbild der Distanzen zu erwarten, wie bei Stresswerten unter 0,1. Besonders oberhalb von 0,15 sollte nicht zu viel Vertrauen in die Details des Plots gesetzt werden. Bei Stresswerten zwischen 0,2 und 0,3 sollte die Ordination mit Skepsis betrachtet werden. Die Distanzen werden in diesem Fall in der Ordination nicht gut wiedergegeben (Clark and Warwick, 2001).

III. Ergebnisse

III.1. Bestimmung von Nukleotidsequenzen und funktionellen Unterschieden der im Kreuzungsexperiment kombinierten *csd*-Allele

III.1.1 Das Kreuzungsexperiment

In dem Kreuzungsexperiment (Abbildung III.1), das von M. K. Fondrk und R.E. Page durchgeführt wurde, wurden insgesamt 15 *csd*-Allele jeweils paarweise miteinander kombiniert. Als Allel wird eine *csd*-Sequenz bezeichnet, die sich durch Nukleotiddifferenzen von anderen *csd*-Sequenzen unterscheidet. Die verschiedenen *csd*-Allele müssen jedoch nicht funktionell verschieden sein, da nicht alle Nukleotiddifferenzen eine Bedeutung für die Spezifität der Allele haben. In der Parental (P) -Generation des Kreuzungsexperimentes wurden fünf P-Königinnen mit Spermien von verschiedenen Drohnen einzelbesamt (insgesamt fünf P-Kreuzungen). Die Spermien der Filialgeneration 1 (F1) -Drohnen einer P-Königin wurden vereinigt (insgesamt fünf Spermienmischungen). Königinnen der F1-Generation wurden in 23 F1-Kreuzungen mit je einem F1-Spermienpool besamt. Werden funktionell gleiche *csd*-Allele oder sogar gleiche *csd*-Sequenzen kombiniert, so stellt dies den homozygoten Zustand dar. Homozygote *csd*-Allele führen zu diploiden Drohnen. Diese werden kurz nach dem Schlupf von Arbeiterinnen gefressen und erreichen nicht das Adultstadium (Mackensen, 1955). Fünf der 23 F1-Kreuzungen waren Inzuchtkreuzungen, in denen F1-Königinnen mit den Spermien der Geschwisterdrohnen besamt wurden (Tabelle III.2). In den Inzuchtkreuzungen bestand jeweils eine der vier Allelkombinationen aus gleichen *csd*-Sequenzen, so dass die Eier der Nachkommen zu 50% homozygot für *csd* waren. In den heterozygoten F2-Arbeiterinnen treten nur zwei funktionell verschiedene *csd*-Allele auf. Daher sollten homozygote *csd*-Allele in den befruchteten F2- Eiern existieren, in den adulten F2-Arbeiterinnen aber fehlen.

In dem Kreuzungsexperiment wurden unterschiedliche *csd*-Allele insgesamt 71mal paarweise kombiniert. Um mögliche Einflüsse der Herkunft der Allele (paternal oder maternal) auf deren Spezifität aufzudecken, wurden bei den Kreuzungen acht paarweise Allelkombinationen auch reziprok erzeugt. Das bedeutet, dass z.B. eine Allelkombination

aus einem Allel A und einem Allel B sowohl mit maternalem Allel A und paternalem Allel B als auch mit paternalem Allel A und maternalem Allel B gebildet wurde.

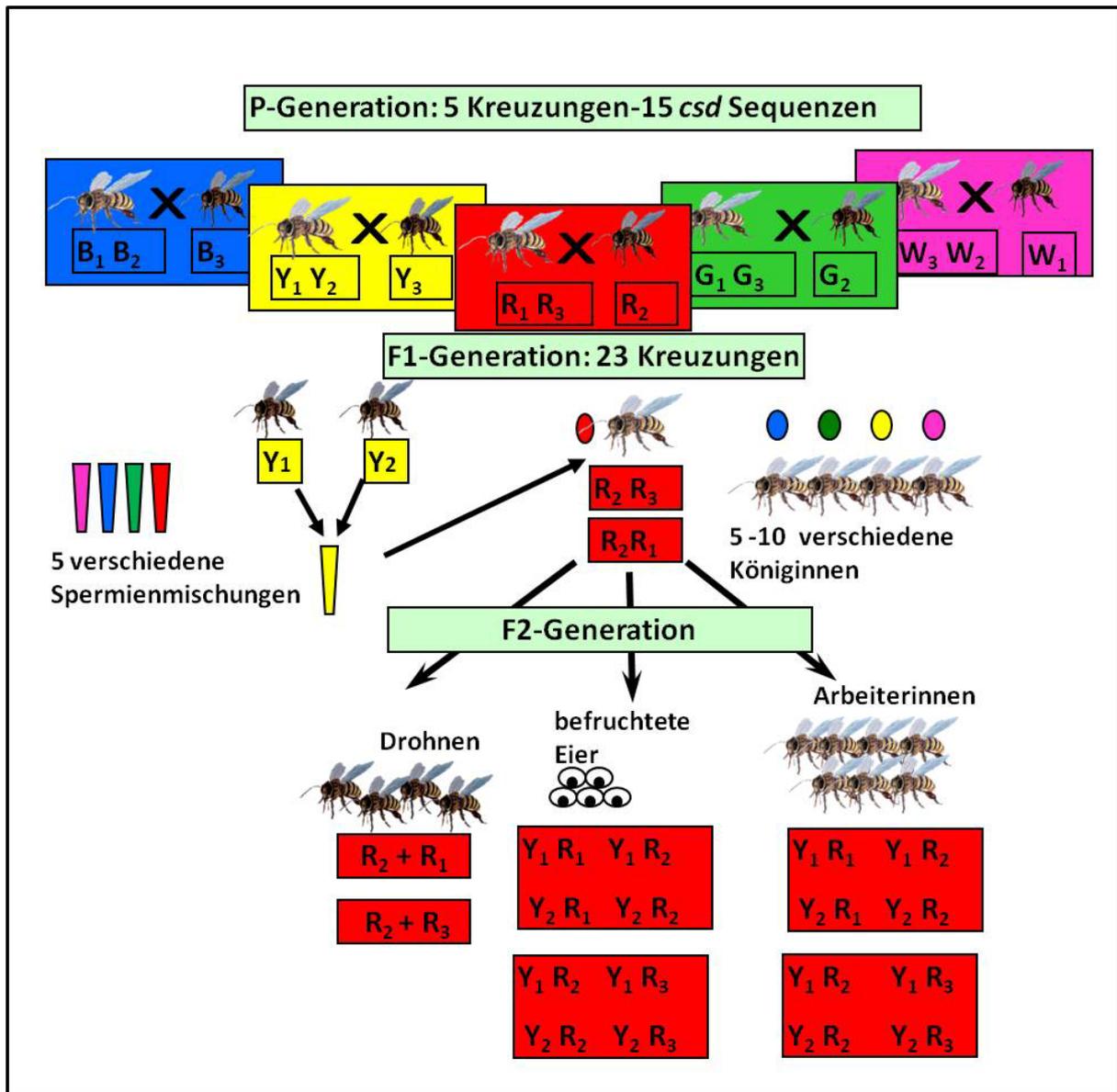


Abbildung III.1: Kreuzungsschema der P- und der F1-Kreuzungen

Die P-Königinnen der verschiedenen Linien (B, G, R, W und Y) werden mit je einem P-Drohn besamt. Die Drohnen der F1-Generation tragen je eines der beiden Allele der P-Königin. Diese beiden Allele befinden sich in den F1-Spermienmischungen der jeweiligen Linien. Für die F1-Königinnen gibt es je Linie zwei mögliche Genotypen. Sie tragen je eines der beiden Allele der P-Königin und das Allel des P-Drohns. Die F2-Drohnen tragen je eines der beiden Allele der F1-Königin. In den befruchteten F2-Eiern existieren vier *csd*-Kombinationen, unabhängig davon, ob sich die kombinierten Allele funktionell unterscheiden. In den F2-Arbeiterinnen treten nur diejenigen *csd*-Kombinationen auf, die aus funktionell verschiedenen *csd*-Allelen bestehen. Dargestellt ist die F1-Kreuzung R8 in der eine F1-Königin der Linie R mit dem Spermienmischung der Linie Y besamt wurde.

III.1.2 Identifikation der *csd*-Genotypen der P- und F1- Generationen und der F2-Drohnen

Die Nukleotidsequenzen der *csd*-Allele der F1- und der F2- Drohnen wurden durch Restriction-Fragment-Length-Polymorphism (RFLP) Analyse der Exon 6-9 Region genomischer DNA identifiziert und die identifizierten Allele kloniert und sequenziert. Verschiedene *csd*-Allele weisen verschiedene Restriktionsmuster auf. Bei der RFLP Analyse von Drohnen entsteht das Restriktionsmuster von nur einem Allel, bei der von weiblichen Bienen von zwei Allelen. Bei Kreuzungen, in denen vier *csd*-Allele kombiniert werden, werden in den in den F2-Arbeiterinnen und den F2-Eiern die gleichen vier *csd*-Allelkombinationen erwartet. Bei den Inzuchtkreuzungen, in denen unter anderem auch zwei gleiche *csd*-Sequenzen kombiniert wurden (homozygoter Zustand), wurden in den F2-Eiern vier verschiedene Allelkombinationen erwartet, in den F2-Arbeiterinnen jedoch nur drei, da hier die homozygoten Allele fehlen sollten. Die Allele der P-Generation und der F1-Königinnen mussten identifiziert werden, um die erwarteten Genotypen der F2-Eier und F2-Arbeiterinnen bestimmen zu können. Dafür wurden die Allele der F1-Drohnen, die die Allele der P-Königin tragen, und die der F2-Drohnen, die die Allele der F1-Königin und damit das Allel des P-Drohns und ein Allel der P-Königin tragen, bestimmt. Die in den F2-Arbeiterinnen erwarteten Genotypen wurden aus den Genotypen der entsprechenden F1-Königinnen und der für die jeweilige Inseminierung verwendeten Spermienmischung bestimmt. In Tabelle III.1 sind die *csd*-Allele der P-Generationen gezeigt. In Tabelle III.2 sind die *csd*-Allele der F1-Königinnen und der Spermienmischungen der F1-Drohnen angegeben, die in den verschiedenen F1-Kreuzungen für die Inseminierungen verwendeten wurden. Daraus wurde auf die zu erwartenden Genotypen der F2-Arbeiterinnen und der F2-Eier rückgeschlossen. Bei den Kreuzungen R06, R11 und W43 konnte jeweils nur ein Allel in den F2-Drohnen identifiziert werden, so dass hier keine Aussage über alle Genotypen der F2-Arbeiterinnen dieser Kreuzungen gemacht werden konnte. Das jeweils zweite Allel wurde erst bei der Genotypisierung der F2-Arbeiterinnen nachgewiesen (siehe Abschnitt III.1.3). Die betreffenden Allele sind in Tabelle III.2 fett gedruckt. Die Genotypen der F1-Königinnen der Kreuzungen G07 und G08 sind gleich. Beide Königinnen wurden mit derselben Spermienmischung inseminiert (Tabelle III.2). Daraus folgt, dass auch die Genotypen der F2-Generationen dieser beiden Kreuzungen gleich sind.

Die Analyse der Genotypen der F2-Arbeiterinnen und F2-Eier wurde daher auf die Kreuzung G07 beschränkt.

Tabelle III.1: *csd*-Allele der P-Generation

Die *csd*-Allele der fünf P-Königinnen und der fünf P-Drohnen wurden durch Identifikation der *csd*-Allele der F1- und F2-Drohnen bestimmt (siehe Text).

P-Königinnen	Allele	P-Drohnen	Allel
B	B1 B2	B	B3
G	G1 G3	G	G2
R	R1 R3	R	R2
W	W2 W3	W	W1
Y	Y1 Y2	Y	Y3

Tabelle III.2: In der F1- Generation und den F2-Drohnen identifizierte *csd* Allele und in den F2-Arbeiterinnen und F2-Eiern zu erwartende Genotypen

Die Genotypen der F1-Königinnen wurden anhand der F2-Drohnen der entsprechenden Kreuzungen ermittelt, die Allele der fünf verschiedenen Spermienmischungen anhand der F1-Drohnen. Die fett gedruckten Allele der Kreuzungen R06, R11 und W43 konnten in den F2-Drohnen nicht nachgewiesen werden. Die in den F2- Arbeiterinnen und F2-Eiern erwarteten Genotypen wurden aus den Genotypen der entsprechenden F1-Königinnen und der für die jeweilige Inseminierung verwendeten Spermienmischung bestimmt. Bei Kreuzungen, in denen vier *csd*-Allele kombiniert wurden, werden vier verschiedene Allelkombinationen in den F2-Arbeiterinnen und den F2-Eiern erwartet. Bei Inzuchtkreuzungen werden in den F2-Eiern vier verschiedenen Allelkombinationen erwartet, in den F2-Arbeiterinnen nur drei. Die bei den fünf Inzuchtkreuzungen (B12, G07/G08, R11, W47 und Y10) erwarteten homozygoten Allele in den F2-Eiern sind kursiv gedruckt. Bei den Kreuzungen R06, R11 und W43 konnte jeweils nur eines der Allele der F1-Königin bestimmt werden. Die nicht bestimmbar Allele wurden später in den F2-Arbeiterinnen nachgewiesen und sind in dieser Tabelle fett gedruckt.

Name der F1-Kreuzung	<i>csd</i> Allele der F1-Königin	<i>csd</i> -Allele der F1-Spermienmischung	erwartete <i>csd</i> -Allele der F2-Arbeiterinnen	erwartete <i>csd</i> -Allele der F2-Eier
B05	B2 B3	G1 G3	B2 G1; B2 G3; B3 G1; B3 G3	B2 G1; B2 G3; B3 G1; B3 G3
B10	B1 B3	R1 R3	B1 R1; B1 R3; B3 R1; B3 R3	B1 R1; B1 R3; B3 R1; B3 R3
B12	B1 B3	B1 B2	B1 B2; B1 B3; B2 B3	B1 B2; B1 B3; B2 B3; <i>B1 B1</i>
G07	G1 G2	G1 G3	G1 G2; G1 G3; G2 G3	G1 G2; G1 G3; G2 G3; <i>G1 G1</i>
G08	G1 G2	G1 G3	G1 G2; G1 G3; G2 G3	G1 G2; G1 G3; G2 G3; <i>G1 G1</i>
G10	G2 G3	Y1 Y2	G2 Y1; G2 Y2; G3 Y1; G3 Y2	G2 Y1; G2 Y2; G3 Y1; G3 Y2
G12	G2 G3	R1 R3	G2 R1; G2 R3; G3 R1; G3 R3	G2 R1; G2 R3; G3 R1; G3 R3
G14	G2 G3	B1 B2	G2 B1; G2 B2; G3 B1; G3 B2	G2 B1; G2 B2; G3 B1; G3 B2
R06	R1 R2	G1 G3	G1 R1 ; G1 R2; G3 R1 ; G3 R2	G1 R1 ; G1 R2; G3 R1 ; G3 R2
R11	R1 R2	R1 R3	R1 R2 ; R1 R3; R2 R3	R1 R2 ; R1 R3; R2 R3; <i>R1 R1</i>
R13	R2 R3	B1 B2	R2 B1; R2 B2; R3 B1; R3 B2	R2 B1; R2 B2; R3 B1; R3 B2
R14	R2 R3	W2 W3	R2 W2; R2 W3; R3 W2; R3 W3	R2 W2; R2 W3; R3 W2; R3 W3
R8	R1 R2	Y1 Y2	R1 Y1; R1 Y2; R2 Y1; R2 Y2	R1 Y1; R1 Y2; R2 Y1; R2 Y2
W40	W1 W3	G1 G3	W1 G1; W1 G3; W3 G1; W3 G3	W1 G1; W1 G3; W3 G1; W3 G3
W41	W1 W2	Y1 Y2	W1 Y1; W1 Y2; W2 Y1; W2 Y2	W1 Y1; W1 Y2; W2 Y1; W2 Y2
W43	W1 W3	R1 R3	W1 R1; W1 R3; W3 R1; W3 R3	W1 R1; W1 R3; W3 R1; W3 R3
W45	W1 W2	B1 B2	W1 B1; W1 B2; W2 B1; W2 B2	W1 B1; W1 B2; W2 B1; W2 B2
W47	W1 W2	W2 W3	W1 W2; W1 W3; W2 W3	W1 W2; W1 W3; W2 W3; W2 W2
Y06	Y2 Y3	G1 G3	Y2 G1; Y2 G3; Y3 G1; Y3 G3	Y2 G1; Y2 G3; Y3 G1; Y3 G3
Y10	Y1 Y3	Y1 Y2	Y1 Y2; Y1 Y3; Y2 Y3	Y1 Y2; Y1 Y3; Y2 Y3; <i>Y1 Y1</i>
Y12	Y1 Y3	R1 R3	Y1 R1; Y1 R3; Y3 R1; Y3 R3	Y1 R1; Y1 R3; Y3 R1; Y3 R3
Y13	Y2 Y3	B1 B2	Y2 B1; Y2 B2; Y3 B1; Y3 B2	Y2 B1; Y2 B2; Y3 B1; Y3 B2
Y15	Y2 Y3	W2 W3	Y2 W2; Y2 W3; Y3 W2; Y3 W3	Y2 W2; Y2 W3; Y3 W2; Y3 W3

III.1.3 Identifikation funktionell verschiedener *csd*-Allele in den Arbeiterinnen der F2-Generation

Es wurden je Kreuzung die *csd*-Allele von 14 bis 45 F2-Arbeiterinnen mittels RFLP bestimmt, um die funktionell verschiedenen *csd*-Allele in den heterozygoten Individuen identifizieren zu können. Dafür wurde ein Primerpaar entworfen mit dessen Hilfe die Exon 6-9 Regionen aller *csd*-Allele auf genomischer Ebene amplifiziert werden konnten. Bei der PCR für die RFLP-Analyse wurden mit [α -P³³]-markierte dATP Moleküle zugegeben, um die PCR-Produkte radioaktiv zu markieren. Nach der Restriktion wurden die DNA-Fragmente über ein hochauflösendes Polyacrylamidgel bis auf Größenunterschiede von wenigen Nukleotiden aufgetrennt. Die radioaktiven Banden wurden mittels Röntgenfilm detektiert. Durch die scharfe Trennung ließen sich die Restriktionsmuster der paarweise kombinierten Allele eindeutig voneinander unterscheiden. In Abbildung III.2 ist die RFLP-Analyse am Beispiel von elf F2-Arbeiterinnen der F1-Kreuzung W40 dargestellt. Die PCR-Produkte wurden mit der Restriktionsendonuklease *SspI* restringiert und gelelektrophoretisch aufgetrennt. Durch Vergleich der Restriktionsmuster der F2-Arbeiterinnen mit den Restriktionsmustern der Einzelallele der F2-Drohnen, konnten die Allelkombinationen in den F2-Arbeiterinnen identifiziert werden.

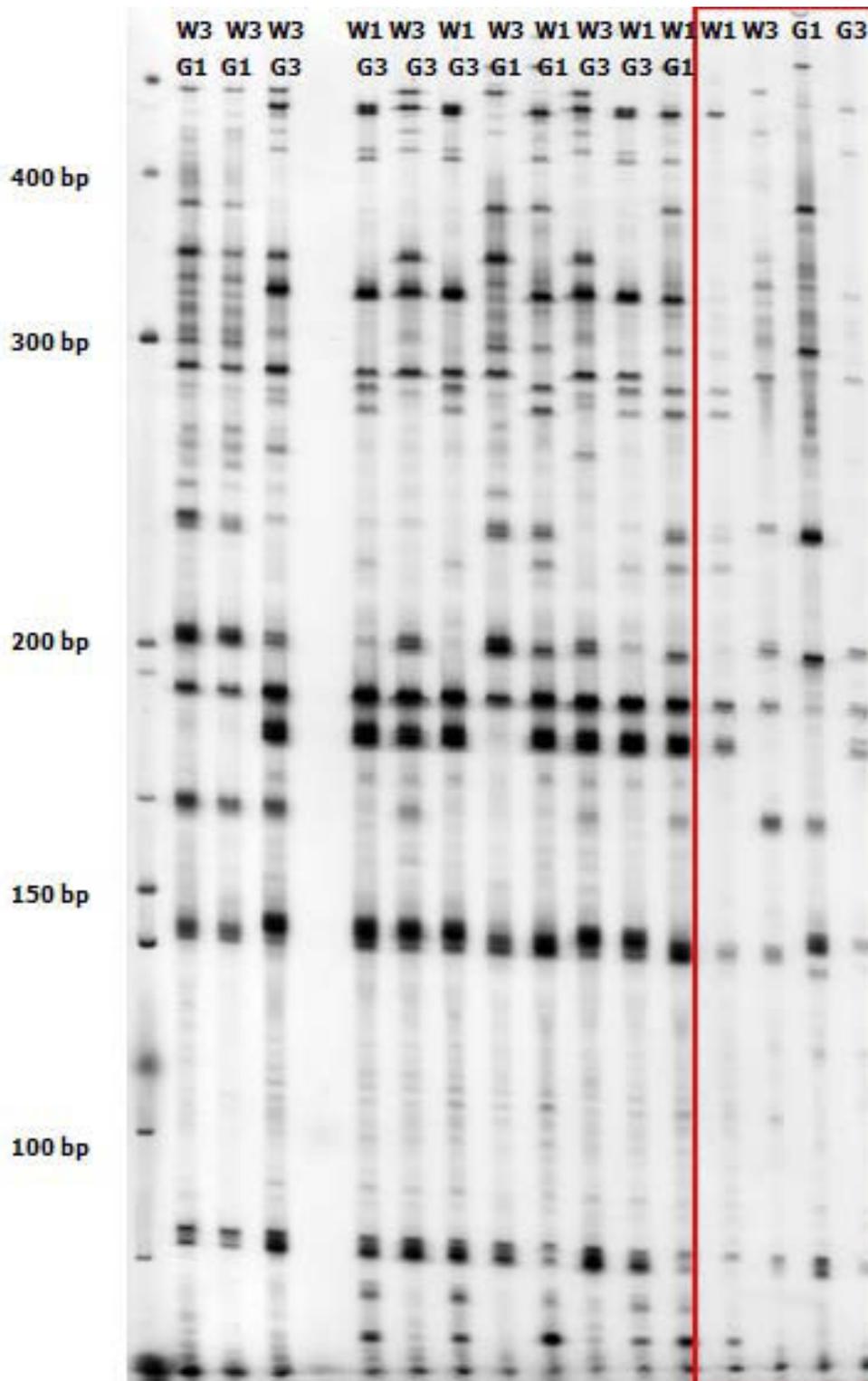


Abbildung III.2: Genotypisierung von F2-Arbeiterinnen. Dargestellt sind die Restriktionsmuster von elf F2-Arbeiterinnen der Kreuzung W40 und vier F2-Drohnen (roter Rahmen), die die in den Arbeiterinnen erwarteten Allele tragen. Die PCR-Reaktionen wurden unter Zugabe von $[\alpha\text{-P}^{33}]$ dATP durchgeführt und die PCR-Produkte mittels der Restriktionsendonuklease *SspI* restringiert. Die Auftrennung erfolgte anhand eines hochauflösenden, denaturierenden Polyacrylamidgels. Die Allelkombinationen der Arbeiterinnen wurden durch Vergleich ihrer Restriktionsmuster mit den Restriktionsmustern der Einzelallele der Drohnen identifiziert. Sie sind am oberen Ende der jeweiligen Spur angegeben.

Unter der Annahme, dass die vier bzw. drei (bei Inzuchtkreuzungen) möglichen *csd*-Allelkombinationen einer Kreuzung gleichhäufig auftreten wäre die Wahrscheinlichkeit (P), dass jede der vier bzw. drei Allelkombinationen in den F2-Arbeiterinnen auftritt $P=1/4$ (0,25) bzw. $P=1/3$ (0,33). Unter dieser Prämisse müssten 16 Individuen getestet werden, um mit 95-%iger Wahrscheinlichkeit jede vorhandene Allelkombination mindestens einmal zu finden. Wird eine mögliche Allelkombination in 16 Arbeiterinnen nicht gefunden, ist die Annahme, dass diese Allelkombination nicht existiert mit 95-%iger Wahrscheinlichkeit richtig. Falls eine Allelkombination nach dem Test von 16 verschiedenen Arbeiterinnen nicht nachgewiesen werden konnte, wurden bis zu 29 weitere F2-Arbeiterinnen getestet (Tabelle III.3), um sicher zu gehen, dass die Allelkombination wirklich nicht in den F2-Arbeiterinnen existierte. Jede gefundene Allelkombination wurde mindestens dreimal unabhängig nachgewiesen. Die Anzahl der möglichen Tests wurde durch die vorhandene Stichprobengröße von 40-45 F2-Arbeiterinnen pro Kreuzung beschränkt. Es wurden bei den Tests alle, in den F2-Arbeiterinnen erwarteten Allelkombinationen (Tabelle III.2) gefunden. Bei Kreuzung Y06 konnte die Allelkombination Y2|G1 nur einmal nachgewiesen werden, da von den zur Verfügung stehenden Arbeiterinnen (n=45) nur eine diese Allelkombination trug. Alle anderen Allelkombinationen wurden je Kreuzung mindestens dreimal bestätigt. Diese Funde zeigen, dass alle Allelkombinationen aus funktionell verschiedenen *csd*-Allelen bestehen. Bei den Inzuchtkreuzungen wurden außer bei den F2-Arbeiterinnen auch bei einzelnen F2-Eiern die *csd*-Allele identifiziert, um zu prüfen, ob die in den Arbeiterinnen fehlenden, homozygoten Allele in den befruchteten Eiern vorhanden sind. Die Anzahl der getesteten Individuen je Kreuzung und die absoluten Häufigkeiten der verschiedenen Allelkombinationen sind in Tabelle III.3 zusammengefasst. In einigen Eiern jeder Inzuchtkreuzung wurde das Restriktionsmuster von jeweils nur einem Allel gefunden. Das gefundene Allel entsprach dabei in jeder Kreuzung den zu erwartenden homozygoten Allelen (Tabelle III.2). Abbildung III.3 gibt einen Überblick über die in den F2-Generationen gefundenen *csd*-Allelkombinationen. Die in den F2-Eiern gefundenen, in den F2-Arbeiterinnen fehlenden, homozygoten Allele sind rot dargestellt. Die acht reziprok erzeugten *csd*-Allelkombinationen sind in Abbildung III.3 grün markiert. In Abbildung III.4 sind die paarweisen Allelkombinationen in Abhängigkeit davon dargestellt, welches der beiden Allele vom Drohnen und welches von der Königin stammt. Die reziprok kombinierten Allelkombinationen sind jeweils in derselben Farbe dargestellt.

Tabelle III.3: Absolute Häufigkeiten der in den F2-Arbeiterinnen und F2-Eiern nachgewiesenen Genotypen. Angegeben sind die gefundenen Allelkombinationen und in Klammern deren absolute Häufigkeit (n_i) sowie die Gesamtzahl n der getesteten Individuen pro Kreuzung. Die homozygoten Allelkombinationen in den F2-Eiern sind kursiv gedruckt.

Name der F1-Kreuzung	<i>csd</i> Allele der F2- Arbeiterinnen				n
	n₁	n₂	n₃	n₄	
B05	B2 G1 (8)	B2 G3 (4)	B3 G1 (7)	B3 G3 (16)	31
B10	B1 R1 (11)	B1 R3 (6)	B3 R1 (3)	B3 R3 (4)	24
B12	B1 B2 (3)	B1 B3 (14)	B2 B3 (9)		26
G07	G1 G2 (9)	G1 G3 (7)	G2 G3 (12)		28
G10	G2 Y1 (10)	G2 Y2 (9)	G3 Y1 (8)	G3 Y2 (3)	30
G12	G2 R1 (6)	G2 R3 (5)	G3 R1 (10)	G3 R3 (3)	24
G14	G2 B1 (8)	G2 B2 (3)	G3 B1 (8)	G3 B2 (5)	24
R06	R1 G1 (9)	R1 G3 (11)	R2 G1 (3)	R2 G3 (5)	28
R08	R1 Y1 (11)	R1 Y2 (3)	R2 Y1 (12)	R2 Y2 (7)	33
R11	R1 R2 (11)	R1 R3 (3)	R2 R3 (9)		23
R13	R2 B1 (11)	R2 B2 (5)	R3 B1 (12)	R3 B2 (6)	34
R14	R2 W2 (4)	R2 W3 (3)	R3 W2 (3)	R3 W3 (4)	14
W40	W1 G1 (5)	W1 G3 (4)	W3 G1 (7)	W3 G3 (4)	20
W41	W1 Y1 (10)	W1 Y2 (4)	W2 Y1 (4)	W2 Y2 (3)	21
W43	W1 R1 (6)	W1 R3 (4)	W3 R1 (6)	W3 R3 (3)	19
W45	W1 B1 (15)	W1 B2 (3)	W2 B1 (5)	W2 B2 (4)	27
W47	W1 W2 (8)	W1 W3 (8)	W2 W3 (10)		26
Y06	Y2 G1 (1)	Y2 G3 (15)	Y3 G1 (11)	Y3 G3 (18)	45
Y10	Y1 Y2 (4)	Y1 Y3 (19)	Y2 Y3 (10)		33
Y12	R1 Y1(13)	R1 Y3 (8)	R3 Y1 (3)	R3 Y3 (8)	32
Y13	Y2 B1 (5)	Y2 B2 (6)	Y3 B1 (7)	Y3 B2 (4)	22
Y15	Y2 W2 (4)	Y2 W3 (4)	Y3 W2 (8)	Y3 W3 (4)	20
	<i>csd</i> Allele der F2-Eier				
B12	<i>B1 B1 (7)</i>	B1 B2 (2)	B1 B3 (3)	B2 B3 (4)	16
G07	<i>G1 G1 (6)</i>	G1 G2 (5)	G1 G3 (4)	G2 G3 (4)	19
R11	R1 R1 (4)	R1 R2 (3)	R1 R3 (2)	R2 R3 (5)	14
W47	W1 W2 (11)	W1 W3 (0)	W2 W2 (5)	W2 W3 (5)	21
Y10	<i>Y1 Y1 (5)</i>	Y1 Y2 (5)	Y1 Y3 (5)	Y2 Y3 (2)	17

	B1	B2	B3	G1	G2	G3	R1	R2	R3	W1	W2	Y1	Y2	Y3
B1	o													
B2	o													
B3	o	o												
G1		o	o	o										
G2	o	o		o										
G3	o	o	o	o	o									
R1	o		o	o	o	o	o							
R2	o	o		o	o	o	o							
R3	o	o	o		o	o	o	o						
W1	o	o		o	o	o	o		o					
W2	o	o			o			o	o	o	o			
Y1					o	o	o	o	o	o	o	o		
Y2	o	o		o	o	o	o	o		o	o	o		
Y3	o	o		o	o	o	o		o		o	o	o	

Abbildung III.3: In den F2-Generationen gefundene *csd*-Allelkombinationen. Experimentell in F2-Arbeiterinnen nachgewiesene *csd* Allelkombinationen sind schwarz bzw. grün (reziprok vorkommende Allelkombinationen) dargestellt. Die homozygoten Allele wurden nur in den F2-Eiern nachgewiesen und sind rot dargestellt.

Königin

	B1	B2	B3	G1	G2	G3	R1	R2	R3	W1	W2	Y1	Y2	Y3
B1	o		o		o	o		o	o	o	o		o	o
B2	o		o		o	o		o	o	o	o		o	o
B3														
G1		o	o	o	o		o	o		o			o	o
G2								o	o	o	o		o	o
G3		o	o	o	o		o	o	o	o			o	o
R1	o		o		o	o	o	o		o		o		o
R2														
R3	o		o		o	o	o	o		o		o		o
W1														
W2								o	o	o	o		o	o
Y1				o	o	o	o	o	o	o	o	o		o
Y2					o	o	o	o		o	o	o		o
Y3											o	o		o

Drohn

Abbildung III.4: In den F2-Generationen gefundene *csd*-Allelkombinationen der Allele nach deren Herkunft. Experimentell in der F2-Generation nachgewiesene einfach vorkommende *csd* Allelkombinationen sind schwarz dargestellt. Die reziprok vorkommenden Allelkombinationen sind jeweils paarweise in der gleichen Farbe markiert. Als Beispiel sei die Allelkombination B2|G3 herausgegriffen. Diese ist in roter Farbe dargestellt. Bei der links angeordneten Allelkombination stammt das Allel G3 von dem Drohnen und das Allel B2 von der Königin. Bei der weiter rechts angeordneten Allelkombination stammt Allel B2 vom Drohnen und Allel G3 von der Königin.

III.1.4 Bestimmung der Aminosäuresequenzen *csd*-Allele

Um die Aminosäuresequenzen der *csd*-Allele ermitteln zu können, mussten zuerst zunächst die Nukleotidsequenzen der Allele in voller Länge bestimmt werden. Für die Bestimmung der Nukleotidsequenzen wurden die Transkripte aus Pools von 40-50 F2-Eiern der fünf Inzuchtkreuzungen, die jeweils drei *csd*-Allele enthalten, isoliert. Die isolierte mRNA wurde mittels reverser Transkriptase in cDNA umgeschrieben. Aus der cDNA sollten nun die Nukleotidsequenzen der einzelnen Allele getrennt amplifiziert, kloniert und sequenziert werden. Dafür wurden für die PCR-Reaktionen reversed Primer benötigt, mit deren Hilfe gezielt jeweils nur eines der im cDNA-Pool enthaltenen Allele amplifiziert werden konnte (Tabelle A.III und Tabelle A.IV). Diese Primer wurden mit Hilfe der genomischen Exon 6-8 Regionen angefertigt. Mit im 5'-UTR liegenden konservierten forward Primern bzw. 5'RACE forward Primern (Tabelle A. II) und den angefertigten reversed Primern wurden die Nukleotidsequenzen vom 5' UTR bis Exon 6 bzw. 7/8 der einzelnen Allele amplifiziert.

Die Sequenzdaten der 5'-UTR bis Exon 6 -8 und der Exon 6-9 Regionen der jeweiligen Allele wurden mit Hilfe des Programms Bioedit in ihrem überlappenden Bereich in einem Alignment angeordnet, so dass die kompletten Nukleotidsequenzen vom 3'- bis zum 5'-UTR der Allele vorlagen. Nach Entfernen der 5'-Region vor dem Startcodon (Exon 2) und der 3'-Region nach dem Stopcodon (Exon 9) aus diesen Nukleotidsequenzen, wurden die resultierenden Aminosäuresequenzen bestimmt. Nach einem Alignment der 15 Aminosäuresequenzen zeigte sich, dass sich von den 15 im Kreuzungsexperiment verwendeten *csd*-Sequenzen nur 14 in ihrer Aminosäuresequenz unterscheiden, also nur 14 *csd*-Allele in der Kreuzung kombiniert wurden. Bei den *csd*-Sequenzen G2 und W3 handelt es sich um das gleiche Allel. Ihre Daten werden bei den Analysen als Allel G2 zusammengefasst.

III.2 Analysen zur phylogenetischen Verwandtschaft der 14 *csd*-Allele

Mit Hilfe von Nukleotidsubstitutionen in Intronen und synonymen Nukleotidsubstitutionen in den Exonen kann bei Allelen eines Gens, das unter balancierender Selektion steht, analysiert werden, wie eng die Allele phylogenetisch verwandt sind, bzw. wie lange sie schon divergiert sind. Je weniger synonyme Substitutionen und Intronsubstitutionen auftreten, desto enger sind die Allele phylogenetisch verwandt. Für die Bestimmung der phylogenetischen Verwandtschaft der 14 Allele, wurden die Nukleotidsubstitutionen in den Intronen der Exon 6-9 Region mit den synonymen Substitutionen des gesamten Open Reading Frame (ORF) zusammengefasst und mittels des Programmes MEGA analysiert. Anhand der auf diesen Substitutionen beruhenden Ähnlichkeit der Nukleotidsequenzen wurde ein Baum rekonstruiert. Der Test auf Phylogenie wurde mittels der Bootstrappmethode durchgeführt. Bootstrappwerte an den Knotenpunkten geben an, wie verlässlich die Darstellung an dem betrachteten Knotenpunkt ist. Ist der Bootstrappwert an einem Knotenpunkt 95 oder höher, wird die Topologie an diesem Punkt als korrekt angenommen. Abbildung III.5 stellt die phylogenetische Verwandtschaft der 14 Allele dar. Die Bootstrappwerte sind kleiner als 68, bis auf den Knotenpunkt der Allele G1 und Y2.

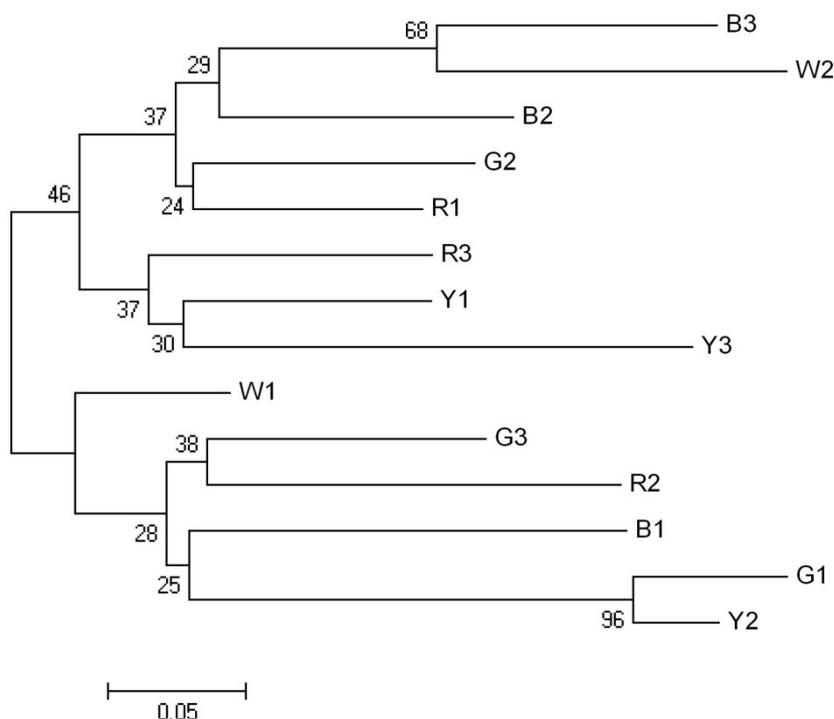


Abbildung III.5: Ähnlichkeitsbaum der Allele basierend auf synonymen Nukleotidsubstitutionen im gesamten *csd*-Gen und Substitutionen in den Intronen der Exon 6-9 Region. Die Topologie des Baumes ist aufgrund der Bootstrappwerte nur am Knotenpunkt der Allele G1 und Y2 gesichert.

Die Topologie des Baumes am Knotenpunkt der Verzweigung G1 – Y2 kann aufgrund des Bootstrapwertes von 96 als gesichert gelten. Diese beiden Allele sind mit großer Wahrscheinlichkeit phylogenetisch eng miteinander verwandt. Die Topologie der restlichen Verzweigungen ist aufgrund niedriger Bootstrapwerte nicht gesichert. Eine gemeinsame Abstammung kann daher, außer bei den Allelen G1 und Y2, nicht festgestellt werden. Dies bedeutet, dass die phylogenetische Verwandtschaft für die Auswertung der Unterschiede in den Aminosäuresequenzen der *csd*-Allele keine Rolle spielt, so dass die Analyse der Aminosäureunterschiede unabhängig von phylogenetischer Verwandtschaft erfolgen kann, außer bei den Allelen G1 und Y2.

III.3 Bearbeitung der Sequenzdaten für die Analyse mit Hilfe von Zwei-Weg hierarchischer Clusteranalyse und Multidimensionaler Skalierung

III.3.1 Transformation der Daten aus den Sequenzvergleichen in Datenmatrices

Die Informationen über die Aminosäureunterschiede der *csd*-Allele in den 71 paarweisen Allelkombinationen werden zunächst in Datenmatrices transformiert. Die Ähnlichkeit der Datenreihen (Allele oder Aminosäurepositionen) aus den Datenmatrices wird durch Berechnung von Ähnlichkeitsmatrices bestimmt. Die Ähnlichkeitsmatrices werden dann mittels Zwei-Weg hierarchischer Clusteranalyse (ZWHC) oder Multidimensionaler Skalierung (MDS) analysiert, um potentiell für Spezifität verantwortliche Aminosäurepositionen identifizieren zu können.

Die Nukleotidsequenzen der 14 Allele in ganzer Länge wurden in Aminosäuresequenzen übersetzt und in einem Alignment angeordnet. Das Alignment enthielt 431 Aminosäurepositionen. Im Bereich der hypervariablen Region (HVR, Aminosäurepositionen 325 bis 351) konnten die Aminosäuren aufgrund von Unterschieden in der Länge der hypervariablen Regionen und Zusammensetzung aus unterschiedlichen Repeats nicht in einem Alignment angeordnet werden. Die HVR wurde daher für die Erstellung der Datenmatrices aus dem Alignment entfernt und später separat ausgewertet (III.7). Das Alignment enthielt daraufhin 404 Aminosäurepositionen. Die Informationen über die Unterschiede der Aminosäuresequenzen in den 71 Allelkombinationen mussten in eine für die Algorithmen der genutzten Programme (Cluster 3.0 bei der ZWHC und Primer 6 bei der MDS) erkennbare Kodierung (Zahlenwerte) transformiert werden.

III.3.1.1 Datenmatrix basierend auf Aminosäureunterschieden: Matrix 1

Für die Datenmatrix wurden die einzelnen Aminosäurepositionen in den 71 experimentell durchgeführten paarweisen Vergleichen danach bewertet, ob sich die Aminosäuren beider Allele an einer Position unterschieden oder nicht. Die Qualität des Unterschieds, d.h. die involvierten Aminosäuren, spielte dabei keine Rolle. Unterschieden sich die Aminosäuren, oder hatte eine der Sequenzen an einer Position eine Aminosäuredeletion, so wurde dieser

Aminosäureposition in dem Vergleich der Wert 1 zugeordnet. Waren die beiden Aminosäuren gleich, bekam die Aminosäureposition den Wert 0. Um nun die für jedes Allel typischen Positionswerte für jede Aminosäureposition zu ermitteln, wurden diejenigen paarweisen Vergleiche, in denen die jeweilige Sequenz vorkam, untereinander geschrieben und der Mittelwert für jede Aminosäureposition errechnet. Jeder Mittelwert entsprach einem für dieses Allel typischen Positionswert und lag zwischen null und eins. Die Mittelwerte enthalten nicht für jedes Allel den gleichen Informationsgehalt, da die verschiedenen Allele nicht in der gleichen Anzahl von experimentellen Allelkombinationen vorkommen. Je höher ein Positionswert war, desto größer war der Unterschied des betrachteten Allels an dieser Position zu den meisten oder allen anderen Allelen. Je niedriger ein Positionswert war, desto weniger unterschied sich das Allel an der betrachteten Stelle von den anderen Allelen. Die für die Allele typische Abfolge von Aminosäuren an den verschiedenen Aminosäurepositionen wurde somit in die für die Allele typische Abfolge der Positionswerte transformiert. Diese Abfolge der Positionswerte an den verschiedenen Aminosäurepositionen wird im Folgenden als alleltypisches Positionsmuster bezeichnet.

Für die Bildung der Matrix wurden die Positionsmuster der 14 *csd*-Allele untereinandergeschrieben. Die Aminosäurepositionen, die bei allen Sequenzen gleich waren (Nullpositionen), wurden gelöscht, da diese für die Spezifität der Allele keine Rolle spielen. Es verblieben 82 Aminosäurepositionen mit Positionswerten, die größer als Null sind. Die entstandene Matrix bestand somit aus 14 Reihen (*csd*-Allelen) und 82 Spalten (Aminosäurepositionen). Diese 14x82-Matrix wird im Folgenden als Matrix 1 (M1) bezeichnet. Ihre Werte sind in Tabelle A.VI angegeben. Jede Reihe gibt das typische Positionsmuster eines Allels an. So wie jedes Allel ein typisches Positionsmuster hat, kann jeder Aminosäureposition eine für diese Position typische Abfolge der Positionswerte angegeben werden, die die 14 Allele von B1 bis Y3 an dieser Aminosäureposition haben. Diese charakteristische Abfolge der Positionswerte ist der Spalte der jeweiligen Aminosäureposition in der Matrix zu entnehmen. Durch Vergleich der Reihen kann die Ähnlichkeit der Allele zueinander analysiert werden und durch Vergleich der Spalten die der Aminosäurepositionen. Matrix 1 ermöglicht Untersuchungen zur Ähnlichkeit von Allelen und von Aminosäurepositionen unter der Annahme, dass alle Aminosäureunterschiede gleichwertig und damit potentiell gleich wichtig für Spezifität von *csd*-Allelen sind.

III.3.1.2 Datenmatrix basierend auf physikochemischen Differenzen der Aminosäureunterschiede: Matrix 2

Eine weitere Möglichkeit zur Bildung einer Matrix besteht darin, die Aminosäureunterschiede in den 71 Allelkombinationen qualitativ zu bewerten, nach der Differenz der physikochemischen Eigenschaften der verglichenen Aminosäuren. Die physikochemischen Differenzen zwischen den 190 möglichen Aminosäurepaaren wurden nach Grantham (1974) berechnet. In die Berechnung gingen Zusammensetzung, Polarität und Volumen der Aminosäure ein. Für die Erstellung der Datenmatrix wurden, wie bei Matrix 1, die Sequenzen der 71 Allelkombinationen paarweise verglichen. Jede Aminosäureposition bekam in jedem paarweisen Vergleich einen Wert für die physikochemische Differenz der beiden Aminosäuren an dieser Position. Die physikochemische Differenz ist eine einheitsfreie Kenngröße. Bei den paarweise verglichenen Aminosäuresequenzen entstanden physikochemische Differenzen zwischen 0 und 196. Die Positionswerte in der Matrix können daher entsprechend variabel sein. Die für jedes Allel typischen Positionswerte wurden dann wie bei M1 ermittelt und unterinandergeschrieben. Einige Allele weisen an bestimmten Positionen Aminosäuredeletionen auf. Die Wirkung von Deletionen auf die Spezifität oder die Funktionsfähigkeit der Csd-Proteine ist nicht bekannt. Die Aminosäurepositionen 144, 211, 307, 358, 359, 362, 372, 373, 417 und 431 weisen bei bestimmten Allelen Deletionen auf. Sie wurden bei den paarweisen Vergleichen der entsprechenden Allele nicht einbezogen und gingen deshalb nicht in Berechnung der entsprechenden Positionswerte ein. Die so entstandene Datenmatrix enthielt 14 Reihen (*csd*-Allele) und 82 Spalten (Aminosäurepositionen) und wird als Matrix 2 (M2) bezeichnet. Ihre Werte sind in Tabelle A.VII angegeben. M2 ermöglicht Untersuchungen zur Ähnlichkeit von Allelen und von Aminosäurepositionen unter der Annahme, dass Aminosäureunterschiede umso wichtiger für die funktionelle Verschiedenheit von Allelen sind, je größer deren physikochemische Differenz und damit die Ausprägung des Unterschieds ist. Es interessiert nicht mehr der bloße Unterschied, wie in Matrix 1, sondern die Qualität des Unterschieds. Matrix 2 kann aus technischen Gründen nur mittels ZWHC aber nicht mittels MDS analysiert werden.

III.3.1.3 Datenmatrix basierend auf physikochemischen Aminosäureunterschieden, unter Ausschluss der Aminosäurepositionen, die Deletionen aufweisen: Matrix 3

Die Darstellung der Aminosäuredeletionen in M2 führt dazu, dass einige Positionswerte in der Matrix fehlen. Eine Matrix mit Fehlstellen kann aus technischen Gründen nicht mittels Multidimensionaler Skalierung analysiert werden. Um die Daten über die physikochemischen Unterschiede dennoch mittels Multidimensionaler Skalierung auswerten zu können, wurden die Aminosäurepositionen, die Deletionen aufweisen, aus Matrix 2 entfernt. Die entstandene Matrix besteht aus 14 Reihen (*csd*-Allelen) und 72 Spalten (Aminosäurepositionen) und wird als Matrix 3 (M3) bezeichnet. Ihre Werte sind in Tabelle A.VIII angegeben. Die entfernten Positionen wurden separat ausgewertet (III.5).

III.3.2 Bestimmung der Ähnlichkeit von Datenreihen durch Erstellen von Ähnlichkeits- bzw. Unähnlichkeitsmatrices

Die Analyse der Datenmatrices mit den Algorithmen der Zwei-Weg hierarchischen Clusteranalyse und der Multidimensionalen Skalierung setzt eine Bestimmung der Ähnlichkeit der zu vergleichenden Daten voraus. Dies kann erfolgen durch Erstellen von Ähnlichkeits- oder Unähnlichkeitsmatrices aus den Datenmatrices. Ähnlichkeitsmatrices bzw. Unähnlichkeitsmatrices geben die paarweisen Ähnlichkeiten der Datenreihen (entweder der 14 Allele oder der 82 bzw. 72 Aminosäurepositionen) an. Ähnlichkeitsmatrices werden häufig mit Hilfe von Korrelation, Unähnlichkeitsmatrices mittels Bildung der Euklidischen Distanzen bestimmt. Bei vielen Analysen wird meist nur eine Möglichkeit der Ähnlichkeitsbestimmung gewählt. Um die erhobenen Daten möglichst umfassend analysieren zu können, verwendete ich zwei verschiedene Methoden: A) die Spearman'sche Rangkorrelation und B) die Bildung der Euklidischen Distanzen. Bei der korrelationsbasierten Methode wird das Ausmaß des linearen Zusammenhangs zweier Datenreihen quantifiziert. Es werden dabei nicht die absoluten Werte korreliert, sondern die ihnen zugewiesenen Ränge. Die resultierende Matrix ist eine Ähnlichkeitsmatrix. Je höher die Korrelationskoeffizienten der Allele oder der Aminosäurepositionen, desto stärker ist deren Korrelation. Bei der Ähnlichkeitsmessung zweier Datenreihen mittels Bildung der Euklidischen Distanzen gehen die absoluten Differenzen der verglichenen Positionswerte

direkt in die Rechnung ein. Die resultierende Matrix ist eine Unähnlichkeitsmatrix, in der die Größe der Euklidischen Distanz die Unähnlichkeit der verglichenen Daten angibt. Die Ähnlichkeitsmatrices und Unähnlichkeitsmatrices der Allele und der Positionen der Matrices 1 und 3 befinden sich im Anhang (Tabellen A.XV bis A.XXII).

III.4 Analysen der (Un)Ähnlichkeitsmatrices mittels Zwei-Weg Hierarchischer Clusteranalyse und Multidimensionaler Skalierung

Mit Hilfe der Zwei-Weg Hierarchischen Clusteranalyse (ZWHC) und der Multidimensionalen Skalierung (MDS) wurden die (Un)Ähnlichkeitsmatrices analysiert und die Ähnlichkeiten der Daten dargestellt. Bei den Zwei-Weg hierarchischen Clusteranalysen, die mit Hilfe des Programmes Cluster 3.0 durchgeführt wurden, wurden die (Un)Ähnlichkeitsmatrices der Aminosäurepositionen und der *csd*-Allele parallel nach deren Ähnlichkeit auf dichotome Weise gruppiert unter Bildung eines Ähnlichkeitsbaumes für die Allele und eines Ähnlichkeitsbaumes für die Aminosäurepositionen. Je ähnlicher Aminosäurepositionen sind, desto weniger Verzweigungen trennen sie voneinander und desto kürzer sind auch die Astlängen. Bei großen Astlängen besteht wenig Ähnlichkeit, auch, wenn die Positionen nur durch eine oder wenige Verzweigungen getrennt sind. Das gleiche gilt für die Darstellung der Allelähnlichkeit. Durch die parallele Clusteranalyse von Aminosäurepositionen und Allelen können Zusammenhänge zwischen Ähnlichkeiten von Aminosäurepositionen und Ähnlichkeiten von Allelen erkannt werden.

Die Darstellungen von Ähnlichkeiten zwischen Allelen oder zwischen Aminosäurepositionen durch Multidimensionale Skalierung (MDS) wurden mit Hilfe des Programmes Primer 6 durchgeführt. Der Zweck einer MDS besteht darin, die Datenpunkte, die die Aminosäurepositionen oder die Allele darstellen, in einem zweidimensionalen Raum so zu repräsentieren, dass die relativen Abstände aller Datenpunkte in derselben Rangordnung sind, wie deren relative Unähnlichkeiten. Je enger zwei Datenpunkte zusammenliegen, desto ähnlicher sind sie. Stresswerte geben an, wie gut die Ordination die Abstände der Datenpunkte aus den (Un)Ähnlichkeitsmatrices wiedergibt. Clark and Warwick (2001) geben Anhaltspunkte für die Bewertung der Stresswerte: Stresswerte unter 0,1 bedeuten demnach, dass die Ordination gut ist. Bei Stresswerten unter 0,2 ist ein weniger exaktes Abbild der Distanzen zu erwarten, wie bei Stresswerten unter 0,1. Besonders oberhalb von 0,15 sollte

nicht zu viel Vertrauen in die Details des Plots gesetzt werden. Bei Stresswerten zwischen 0,2 und 0,3 sollte die Ordination mit Skepsis betrachtet werden. Die Distanzen werden in der Ordination nicht gut wiedergegeben (Clark and Warwick, 2001).

Anders als bei der Zwei-Weg hierarchischen Clusteranalyse ist das Verfahren der Multidimensionalen Skalierung nicht dichotom, sondern erfolgt gleichberechtigt anhand der Werte. Es ist allerdings nicht erkennbar, worin die Ähnlichkeiten bzw. die Unterschiede zwischen den Aminosäurepositionen oder den Allelen bestehen, da die Analysen der Aminosäurepositionen und die der Allele nicht parallel durchgeführt werden können. Daher ist z.B. nicht ersichtlich, welche Positionen dazu führen, dass Allele als ähnlich bzw. nicht ähnlich dargestellt werden.

III.4.1 Analysen der Matrix 1

III.4.1.1 Zwei- Weg Hierarchische Clusteranalyse von Matrix 1

III.4.1.1.1 Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher Rangkorrelation

In Abbildung III.6 ist die Zwei-Weg hierarchische Clusteranalyse der Ähnlichkeitsmatrix 1, die auf Spearman'scher Rangkorrelation der Daten aus Matrix 1 beruht, gezeigt. Dargestellt wird das Ergebnis der Analyse mit Hilfe eines Ähnlichkeitsbaums für die Allele und eines Ähnlichkeitsbaums für die Aminosäurepositionen. Diese beiden Ähnlichkeitsbäume sind über ein Rechteck aus quadratisch dargestellten Datenpunkten verbunden. Jeder Datenpunkt entspricht dem Positionswert einer bestimmten Aminosäureposition in einem bestimmten Allel. Je intensiver der gelbe Farbton des Datenpunktes, desto höher der entsprechende Positionswert in der Matrix 1. Je mehr der Farbton ins Schwarze geht, desto niedriger der entsprechende Positionswert. Da in dem Kreuzungsexperiment nicht alle möglichen Allelkombinationen erzeugt wurden, kann es passieren, dass Positionswerte von null in Matrix 1 auftreten. Dies ist der Fall, wenn ein Allel an einer Aminosäureposition als einziges Allel eine Aminosäuresubstitution aufweist, und nicht alle anderen Allele mit diesem Allel experimentell kombiniert wurden. Bei diesen nicht kombinierten Allelen beträgt der Positionswert dann null. Ist der Positionswert null, so ist der Datenpunkt ganz schwarz. Das

Gesamtbild gibt mit Hilfe der Ähnlichkeitsbäume Informationen darüber, wie ähnlich sich Allele bzw. Positionen sind. Gleichzeitig ist anhand der Farbe der Datenpunkte erkennbar, welche Aminosäurepositionen dazu führen, dass Allele sich ähnlich bzw. nicht ähnlich sind. Außerdem ist zu sehen, wie variabel Positionen sind, d.h., ob Aminosäurepositionen bei mehreren Allelen gleichzeitig oder nur bei einzelnen Allelen verändert sind.

Weisen Aminosäurepositionen nur bei einem einen intensiv gelben Farbton auf und beim Rest der Allele einen nahezu schwarzen oder schwarz-gelben Farbton, so bedeutet das, dass sich die Aminosäure, die sich bei den entsprechenden Allelen an dieser Position befindet, von denen der übrigen Allelen unterscheidet. Deren Aminosäuren unterscheiden sich an dieser Position untereinander nicht. Solche Aminosäurepositionen können als einzigartig für die entsprechenden Allele bezeichnet werden.

Es gibt auch Positionen, die bei zwei bis vier Allelen im Vergleich zu den anderen Allelen verändert sind. Diese Positionen werden als charakteristisch für die jeweiligen Allele bezeichnet. Und es gibt Positionen, bei denen sich mehr als vier Allele (bei mehr als ca. ein Drittel der Allele) vom Rest der Allele unterscheiden. Bei diesen trägt meist 30-50% der Allele eine andere Aminosäure als die übrigen 50-70% der Allele. Diese Positionen werden als allgemeinvariabel bezeichnet. Der Wert von vier Allelen wurde von mir festgelegt. Ich sah es als sinnvoll an, zwischen Aminosäurepositionen zu unterscheiden, die bei wenigen Allelen (zwei bis vier) substituiert sind, und Aminosäurepositionen, die stärker variabel sind (bei mehr als vier Allelen substituiert). Ein Drittel der Allele schien mir ein passender Grenzwert für eine solche Unterscheidung zu sein. Alle Aminosäurepositionen und die entsprechenden Merkmale (einzigartig, charakteristisch oder allgemeinvariabel) sind in Tabelle III.4 dargestellt.

Tabelle III.4: Aminosäurepositionen aus Matrix 1. Dargestellt sind die Merkmale der Aminosäurepositionen (einzigartig =E; charakteristisch =C; allgemeinvariabel =A). Gegebenenfalls ist in Klammern angegeben, für welche Allele die Positionen einzigartig bzw. charakteristisch sind.

Position	Merkmal	Position	Merkmal
7	A	307	E (B2)
23	A	308	E (G3)
24	A	310	C (B3, R2, W2, Y1)
42	A	312	C (B2, R1, W1)
54	A	321	A
57	B2	322	A
83	R2	323	A
93	C (G1, G3, R2)	324	C (R1, R2)
104	Y3	352	E (W1)
144	E (R1)	353	E (G3)
162	E (Y3)	356	E (B1)
173	C (Y1, Y3)	358	E (B1)
190	A	359	C (B1, B3, R3, W1)
201	A	362	C (B1, B3, R3, W1)
205	E (Y3)	368	C (W2, Y1, Y3)
211	E (W2)	370	E (B2)
219	A	372	C (R2, W1, W2, Y1)
221	E (W2)	373	C (R2, W1, W2, Y1)
230	C (G3, W2)	374	A
237	E (W1)	376	C (G3, R2, W2, Y1)
246	A	380	E (W2)
247	A	383	C (G1, Y2)
251	C (B1, W2)	385	C (G1, Y2)
256	E (W2)	387	E (G3)
259	E (W2)	388	C (G3, W1, Y2)
265	E (B2)	390	A
267	E (R2)	393	E (Y3)
269	E (G3)	399	A
270	C (G1, R2, Y2)	403	A
273	E (Y3)	405	E (G3)
281	C (B3, G2, R1, W2)	411	B1, W1, Y3
282	E (G3)	417	E (B1)
287	A	418	E (B1)
288	A	419	C (B1, Y1)
293	A	421	E (G3)
295	C (B3, R1, R3)	422	E (R1)
297	C (B2, R3)	427	C (B2, W2)
302	E (G3)	428	C (G2, G3)
303	E (R1)	429	E (G2)
304	C (B2, G2, R1, W1)	430	E (W1)
305	E (B1)	431	C (W1, B2)

Die Positionen können als Einzelpositionen auftreten oder als Kombination von Positionen, die für das / die gleiche(n) Allel(e) einzigartig oder charakteristisch sind oder die bei allen Allelen die gleiche allgemeine Variabilität aufweisen.

Eine einzigartige Einzelposition ist die Position 429. Ihr Farbton ist in Abbildung III.6 bei Allel G2 sehr intensiv und bei den übrigen Allelen fast schwarz. Das Allel G2 unterscheidet sich also an Position 429 von allen anderen Allelen. Diese Position ist für G2 einzigartig. Position 388 ist ein Beispiel für eine Einzelposition, die für drei Allele (G3, W1 und Y2) charakteristisch ist und Position 399 ist ein Beispiel für eine allgemeinvariable Einzelposition.

Bei Aminosäurepositionen einer Kombination wird im Ähnlichkeitsbaum der Aminosäurepositionen der Terminus des entsprechenden Astes als senkrechter Strich, an dem sich die jeweiligen Positionen befinden, dargestellt und es ist erkennbar, bei welchem/welchen Allel(en) die Positionen verändert sind. In Tabelle III.5 sind alle Kombinationen von Positionen angegeben. Insgesamt wurden 12 Kombinationen von Positionen gebildet. Davon sind acht Kombinationen einzigartig für ein Allel, zwei sind charakteristisch für zwei Allele und zwei sind allgemeinvariabel.

Tabelle III.5: Kombinationen von Aminosäurepositionen nach ZWHC von Matrix 1. Unter „Merkmal“ ist dargestellt, ob die Kombination einzigartige (E), charakteristische (C) oder allgemeinvariable (A) Positionen enthält. Gegebenenfalls ist in Klammern angegeben, für welche Allele die Positionen einzigartig bzw. charakteristisch sind.

Nummer der Kombination	Positionen der Kombination	Merkmal
1	7, 23, 24, 54	A
2	144, 303, 422	E (R1)
3	269, 282, 302, 308, 353, 387, 405, 421	E (G3)
4	383, 385	C (G1, Y2)
5	237, 352, 430	E (W1)
6	246, 247	A
7	83, 267	E (R2)
8	104, 162, 205, 273, 393	E (Y3)
9	211, 221, 256, 259, 380	E (W2)
10	57, 265, 307, 370	E (B2)
11	305, 356, 358, 417, 418	E (B1)
12	251, 431	C (B1, W2)

Wie in Tabelle III.5 zu sehen, gibt es für die Allele B1, B2 G3, R1, R2, W1, W2 und Y3 Kombinationen von Aminosäurepositionen, die für das jeweilige Allel einzigartig sind. Das Allel G2 hat eine einzigartige Einzelposition (Tabelle III.4). Die Allele B3, G1, R3, Y1 und Y2 haben gar keine einzigartigen Aminosäurepositionen. Sie haben einige charakteristische Positionen (Tabelle III.4 und III.5). Für die Allele G1 und Y2 und die Allele B1 und W2 gibt es

zum Beispiel je eine Kombination von Aminosäurepositionen, die für beide Allele charakteristisch ist.

Der Ähnlichkeitsbaum der Aminosäurepositionen in Abbildung III.6 ist in zwei Hauptäste unterteilt, die jeweils sowohl allgemeinvariable Einzelpositionen und Kombinationen als auch einzigartige und charakteristische Einzelpositionen und Kombinationen enthalten. Es ist erkennbar, dass allgemeinvariable Einzelpositionen oft mit zwei bis drei weiteren allgemeinvariablen Einzelpositionen zusammen in einem Cluster angeordnet werden. Als Cluster wird bezeichnet, wenn sich zwei oder mehrere Positionen (im Ähnlichkeitsbaum der Positionen) oder Allele (im Ähnlichkeitsbaum der Allele) in Zweigen des Baumes befinden, die von einem Knotenpunkt ausgehen, also unmittelbar benachbart sind. Dies ist z.B. der Fall bei den Positionen 219, 293 und 323. Diese Einzelpositionen besitzen ähnliche allgemeine Variabilität, wie in Abbildung III.6 erkennbar. Cluster allgemeinvariabler Positionen, die ähnliche allgemeine Variabilität aufweisen, sind über den ganzen Ähnlichkeitsbaum verteilt. Daher sind manche allgemeinvariablen Positionen bestimmten einzigartigen Positionen, die sich im gleichen Ast befinden, ähnlicher als anderen allgemeinvariablen Positionen im anderen Ast. Ein Beispiel ist die allgemeinvariable Position 399. Sie ist den einzigartigen Positionen 83 und 267 im Ähnlichkeitsbaum direkt benachbart, also sehr ähnlich. Der allgemeinvariablen Position 201, die sich im anderen Hauptast befindet, ist sie dagegen sehr unähnlich.

Am Ähnlichkeitsbaum der Allele in Abbildung III.6 ist zu erkennen, dass sich die Allele G1, Y2 und G2, die Allele R1 und B3 und die Allele Y1 und R3 jeweils ein Cluster bilden, sich also sehr ähnlich sind. Am ähnlichsten sind dabei die Allele G1 und Y2. Die Allele W2, G3 und B1 sind sowohl untereinander als auch zu den übrigen Allelen am wenigsten ähnlich.

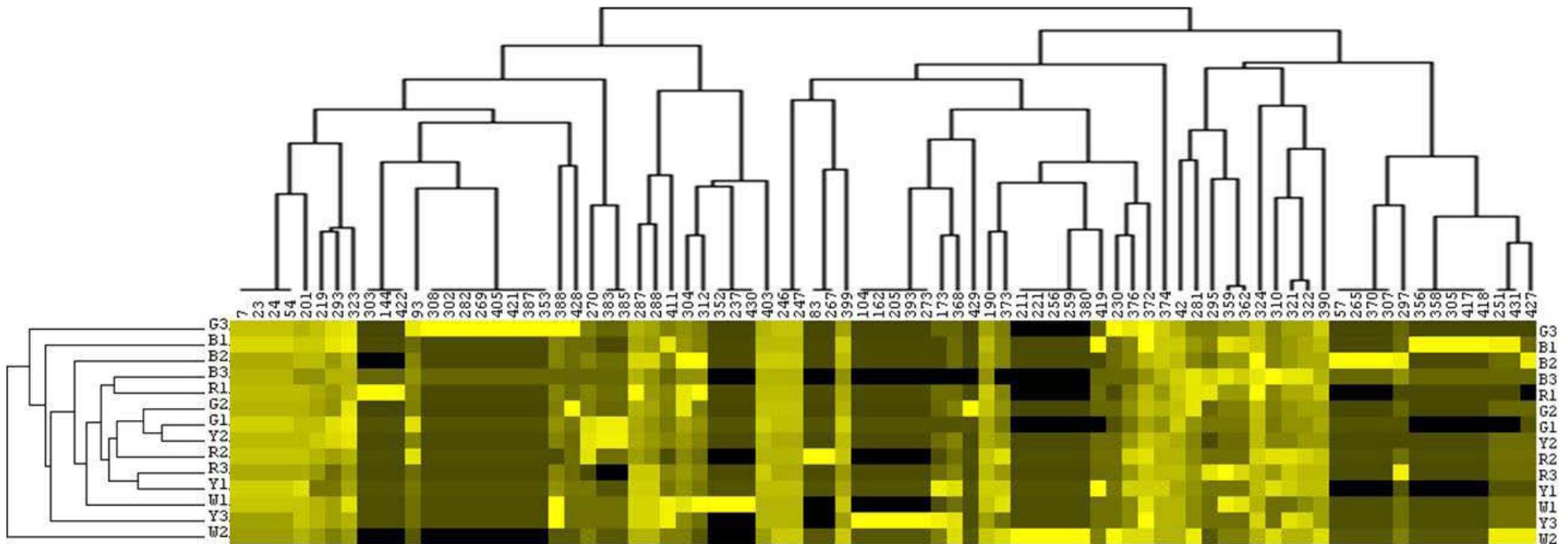


Abbildung III.6: Zwei-Weg Hierarchische Clusteranalyse von Matrix 1. Erstellung der Ähnlichkeitsmatrix erfolgte mittels Spearman'scher Rangkorrelation. Der Ähnlichkeitsbaum der Allele und der der Aminosäurepositionen sind über die graphische Darstellung der Positionswerte in einem Rechteck verbunden. Die Positionswerte aus Matrix 1 werden durch die Datenpunkte dargestellt. Jede Aminosäureposition in einem Allel entspricht einem Datenpunkt, dargestellt als Quadrat. Die Höhe der Positionswerte der Aminosäurepositionen wird durch die Farbintensität ausgedrückt. Je intensiver gelb das Quadrat, desto höher der Positionswert. Je mehr der Farbton ins Schwarze geht, desto niedriger ist der entsprechende Positionswert. Ist ein Datenpunkt ganz schwarz, so wurde das entsprechende Allel experimentell nicht mit dem Allel verglichen, welches an der betrachteten Aminosäureposition eine einzigartige Substitution aufweist.

III.4.1.1.2. Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen

In Abbildung III.7 ist die Clusteranalyse der Unähnlichkeitsmatrix gezeigt, die auf Bildung der Euklidischen Distanzen der Daten aus Matrix 1 beruht. Es werden die gleichen zwölf Kombinationen von Positionen einzigartiger, charakteristischer und allgemeinvariabler Positionen gebildet, wie bei der Analyse der Ähnlichkeitsmatrix, die auf der Spearman'schen Rangkorrelation basiert (Tabelle III.5). Der Ähnlichkeitsbaum der Positionen ist in zwei Hauptäste unterteilt. Die einzigartigen Aminosäurepositionen werden dabei von den allgemeinvariablen Aminosäurepositionen getrennt. In dem einen Hauptast befinden sich alle allgemeinvariablen Einzelpositionen und Kombinationen allgemeinvariabler Positionen, alle charakteristischen Einzelpositionen, die jeweils bei vier Allelen charakteristisch verändert sind, einige charakteristische Einzelpositionen, die bei jeweils drei Allelen verändert sind (Aminosäurepositionen 93, 270, 312, 388, 411) und eine Kombination von charakteristischen Positionen (383 und 385), die bei zwei Allelen verändert sind. In dem anderen Hauptast befinden sich nur einzigartige und charakteristische Einzelpositionen und Kombinationen einzigartiger und charakteristischer Aminosäurepositionen.

Die Allele G1, Y2 und G2, die Allele B3 und R3 und die Allele R2 und Y1 sind sich jeweils sehr ähnlich. Am ähnlichsten sind die Allele G1 und Y2. Am wenigsten ähnlich sind die Allele W2, G3 und B1, sowohl untereinander als auch zu den anderen Allelen.

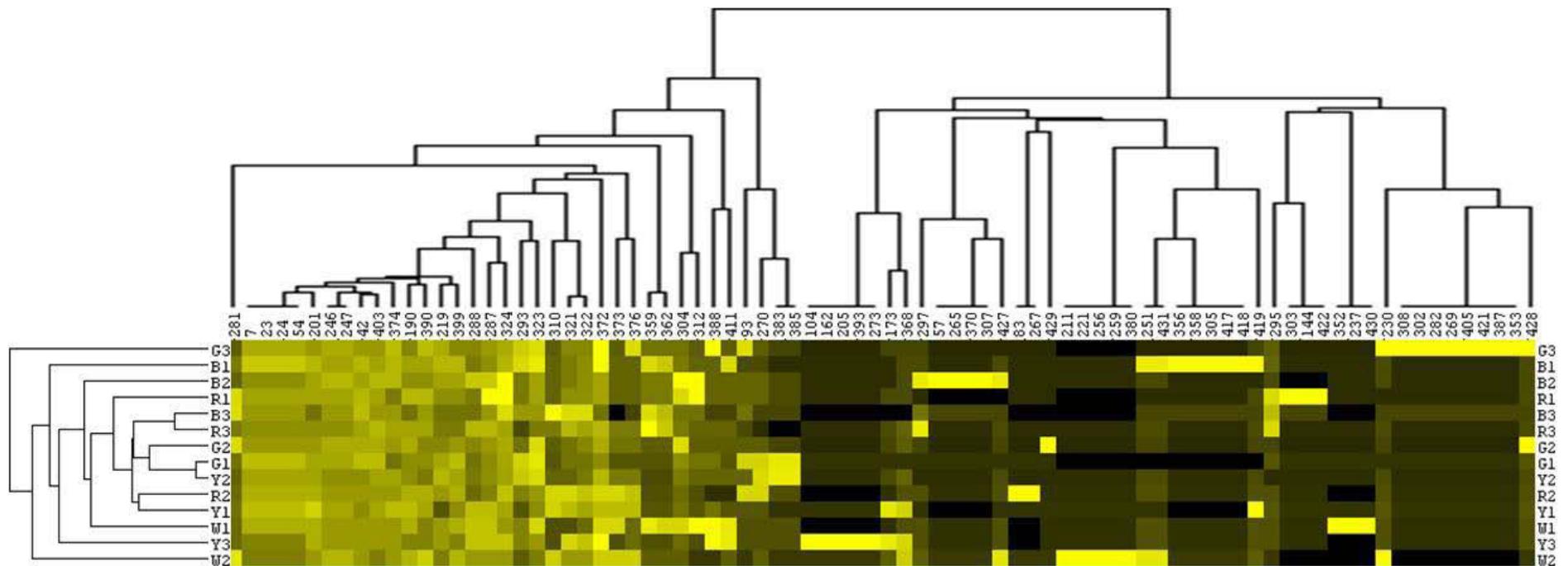


Abbildung III.7: Zwei-Weg Hierarchische Clusteranalyse von Matrix 1. Erstellung der Ähnlichkeitsmatrix erfolgte mittels Bildung der Euklidischen Distanzen. Der Ähnlichkeitsbaum der Allele und der der Aminosäurepositionen sind über die graphische Darstellung der Positionswerte in einem Rechteck verbunden. Die Positionswerte aus Matrix 1 werden durch die Datenpunkte dargestellt. Jede Aminosäureposition in einem Allel entspricht einem Datenpunkt, dargestellt als Quadrat. Die Höhe der Positionswerte der Aminosäurepositionen wird durch die Farbintensität ausgedrückt. Je intensiver gelb das Quadrat, desto höher der Positionswert. Je mehr der Farbton ins Schwarze geht, desto niedriger ist der entsprechende Positionswert. Ist ein Datenpunkt ganz schwarz, so wurde das entsprechende Allel experimentell nicht mit dem Allel verglichen, welches an der betrachteten Aminosäureposition eine einzigartige Substitution aufweist.

III.4.1.2 Multidimensionale Skalierung (MDS) der Allele von Matrix 1

III.4.1.2.1 Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher Rangkorrelation

In Abbildung III.8.A ist die Multidimensionale Skalierung der Allele der Ähnlichkeitsmatrix gezeigt, die auf Spearman'scher Rangkorrelation der Daten aus Matrix 1 beruht. Das Ergebnis der Analyse wird als Graphik dargestellt, in der die Datenpunkte, die die Allele darstellen, in einem zweidimensionalen Koordinatensystem so angeordnet sind, dass die Distanzen der Datenpunkte den Ähnlichkeiten der Allele laut Ähnlichkeitsmatrix möglichst gut entsprechen. Der Stresswert der Graphik gibt an, wie gut die Ähnlichkeiten laut Ähnlichkeitsmatrix wiedergegeben werden. Der Stresswert von Abbildung III.8.A liegt bei 0,18, so dass von einer nicht guten Ordination ausgegangen werden kann, die eine gerade noch zuverlässige, aber nicht detailgetreue Darstellung der Ähnlichkeit der Allele zueinander zeigt. Je geringer der Abstand zweier Datenpunkte ist, desto ähnlicher sind die entsprechenden Allele, die durch die beiden Datenpunkte dargestellt werden. Die meisten Datenpunkte sind in ähnlichen Abständen zueinander um das Zentrum der Graphik angeordnet. Einige Datenpunkte werden in größerem Abstand zum Zentrum in den Randgebieten der Graphik abgebildet. Im Zentrum der Graphik befinden sich die Datenpunkte der Allele G2, R3 und Y2. Diese Datenpunkte haben die kleinsten Abstände zueinander. Die Allele G2, R3 und Y2 sind sich somit untereinander am ähnlichsten. Ähnlichkeit bedeutet in diesem Zusammenhang, dass sich die Ränge, die den Positionen der Allele zugewiesen wurden, nur wenig unterscheiden. Die Datenpunkte der Allele B2, B3, G1, R1, R2, W1 und Y1 befinden sich in geringer Entfernung um dieses Zentrum herum. Diese Allele sind sich untereinander weniger ähnlich als die Allele G2, R3 und Y2. Das bedeutet, dass sich die Rangordnungen der Positionen bei diesen Allelen unähnlich sind. Die Datenpunkte der Allele B1, G3, W2 und Y3 sind werden in den Randgebieten, distanziert vom Zentrum abgebildet. Diese Allele weisen damit zu den jeweils anderen 13 Allelen sehr geringe Ähnlichkeit auf. Die genaue Lage der Datenpunkte der Allele im Koordinatensystem ist in Tabelle A.IX dargestellt.

III.4.1.2.2 Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen

In Abbildung III.8.B ist die Multidimensionale Skalierung der Allele der Unähnlichkeitsmatrix, die auf Bildung der Euklidischen Distanzen der Allele aus Matrix 1 beruht, gezeigt. Der Stresswert liegt bei 0,14. Die Darstellung Ähnlichkeit der Allele ist daher nicht mehr als gut zu bezeichnen. Die Größenordnung der Abstände der Datenpunkte zueinander sollte verlässlich sein, die Detaildarstellung jedoch nicht. Es ist zu erkennen, dass die Datenpunkte der Allele B3, G1, G2, R2, R3, Y1 und Y2 im Zentrum sehr geringe Abstände zueinander haben. Diese Allele sind sich somit sehr ähnlich. Ähnlichkeit bedeutet in diesem Zusammenhang, dass die Euklidische Distanz der Positionen der Allele gering ist. Dies ist der Fall, wenn Allele wenige einzigartige oder charakteristische Positionen aufweisen. In geringer Entfernung zu dieser zentralen Anordnung liegt der Datenpunkt des Allels R1. Die Datenpunkte der Allele B1, B2, G3, W1, W2 und Y3 liegen deutlich distanziert vom Zentrum und voneinander in den Randgebieten. Diese Allele sind somit den jeweils anderen 13 Allelen unähnlich. Das bedeutet, dass diese Allele viele einzigartige und / oder charakteristische Positionen aufweisen, so dass die Euklidischen Distanzen groß werden. Die genaue Lage der Datenpunkte der Allele im Koordinatensystem ist in Tabelle A.X angegeben.

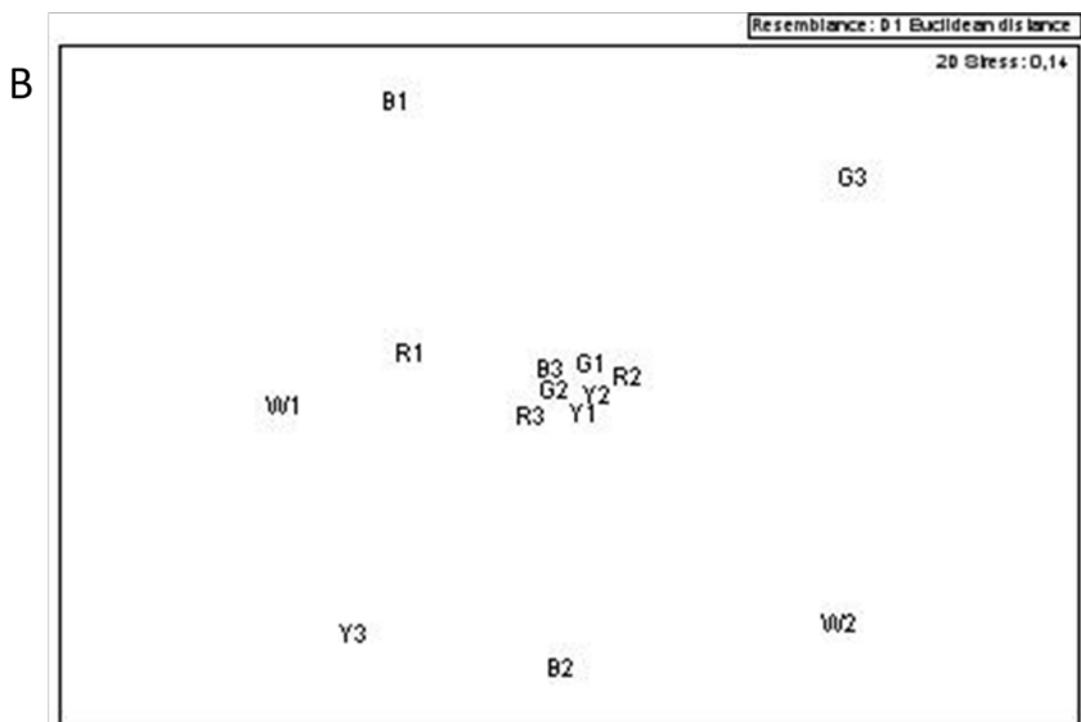
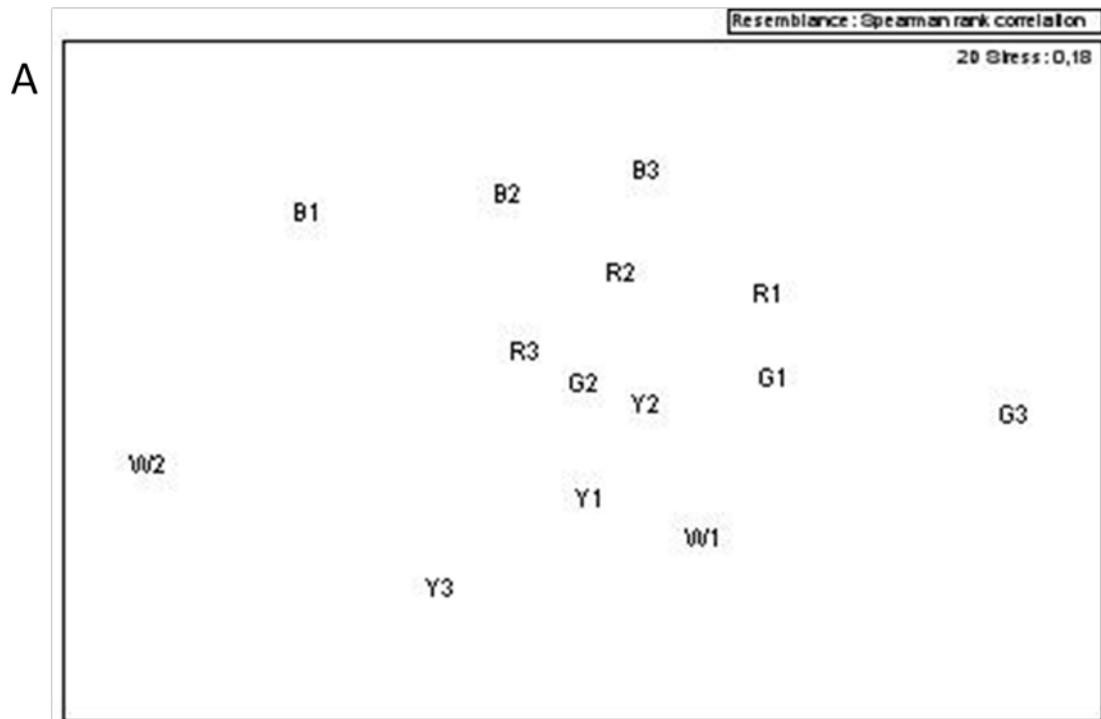


Abbildung III.8: Multidimensionale Skalierung der Allele aus Matrix 1. Berechnung der (Un)Ähnlichkeitsmatrix mittels Spearman'scher Rangkorrelation (A) und Bildung der Euklidischen Distanzen (B) der Daten aus Matrix 1. Je geringer der Abstand der Datenpunkte zweier Allele ist, desto ähnlicher sind diese Allele einander.

III.4.1.3 Multidimensionale Skalierung der Aminosäurepositionen von Matrix 1

III.4.1.3.1 Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher Rangkorrelation

In Abbildung III.9.A ist die Multidimensionale Skalierung der Aminosäurepositionen der Ähnlichkeitsmatrix, die auf Spearman'scher Rangkorrelation der Daten aus Matrix 1 beruht, gezeigt. Der Stresswert der Ordination ist mit 0,26 so hoch, dass die Darstellung der Ähnlichkeiten der Aminosäurepositionen als sehr ungenau angesehen werden kann. Die Graphik wird daher nicht in die Analyse der Ähnlichkeit der Aminosäurepositionen einbezogen.

III.4.1.3.2 Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen

In Abbildung III.9.A ist die Multidimensionale Skalierung der Aminosäurepositionen der Unähnlichkeitsmatrix, die auf Bildung der Euklidischen Distanzen der Daten aus Matrix 1 beruht, gezeigt. Der Stresswert beträgt 0,19. Es ist daher nur eine grobe und keine detailgetreue Darstellung der Ähnlichkeit der Positionen gegeben. Die Datenpunkte der Aminosäurepositionen einer Kombination haben identische Koordinatenwerte. Daher sind sie in der Graphik nicht identifizierbar, können aber anhand ihrer Koordinatenwerte (Anhang Tabelle A.XI) erkannt werden. Die bei dieser Analyse gebildeten Kombinationen von Positionen stimmen mit den Kombinationen überein, die auch unter III.4.1.3.1 und bei der Zwei-Weg hierarchischen Clusteranalyse ermittelt wurden (Tabelle III.5). Die Datenpunkte der Kombinationen allgemeinvariabler Positionen und Einzelpositionen (allgemeinvariabel und charakteristisch) sind in ähnlichen Abständen zueinander auf der einen Seite der Graphik angeordnet. Die Datenpunkte der einzigartigen und charakteristischen Positionen, die sich in Kombinationen befinden und der Datenpunkt der einzigartigen Einzelposition 429 werden davon distanziert am entgegengesetzten Rand der Graphik in geringem Abstand zueinander angeordnet. Diese Kombinationen von einzigartigen und charakteristischen Positionen und die einzigartige Einzelposition sind sich daher untereinander ähnlicher als zu den Kombinationen allgemeinvariabler Positionen und allgemeinvariablen und charakteristischen Einzelpositionen. Diese Ähnlichkeit ergibt sich, weil sich die

Kombinationen von einzigartigen und charakteristischen Positionen und die einzigartige Einzelposition immer nur in einem oder zwei Positionswerten (bei einem Allel oder zwei Allelen) voneinander unterscheiden und die Euklidischen Distanzen deshalb gering bleiben. Die Datenpunkte der Kombinationen der allgemeinvariablen Positionen 246 und 247 und der allgemeinvariablen Positionen 7, 23, 24 und 54 werden mit geringem Abstand zueinander inmitten der Datenpunkte der Einzelpositionen dargestellt (gestrichelter Kreis in Abbildung III.9 (B)). Die Abstände der Datenpunkte der allgemeinvariablen Positionen sind relativ homogen. Da mit einem Stresswert von 0,19 keine detailgetreue Darstellung der Abstände der Datenpunkte gegeben ist, kann anhand der Graphik keine besondere Ähnlichkeit zwischen mehreren allgemeinvariablen Positionen identifiziert werden. Einige Datenpunkte von Einzelpositionen befinden sich ebenfalls in dem gestrichelten Kreis. Mit Hilfe der Koordinatenwerte wurden sie als die Datenpunkte der allgemeinvariablen Positionen 42, 190, 201, 374, 390, 399 und 403 identifiziert.

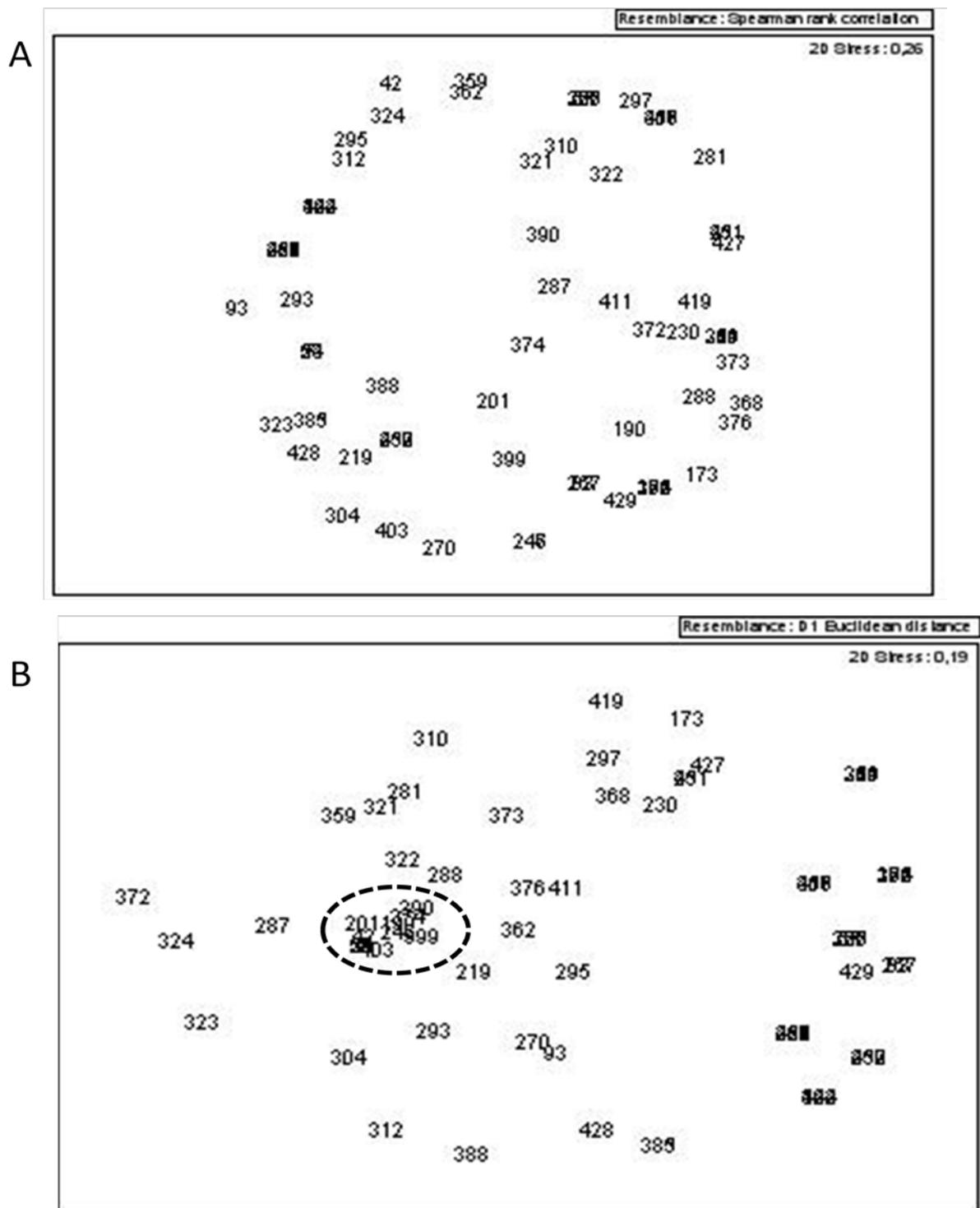


Abbildung III.9: Multidimensionale Skalierung der Aminosäurepositionen aus Matrix 1. Berechnung der (Un)Ähnlichkeitsmatrix mittels Spearman'scher Rangkorrelation (A) und Bildung der Euklidischen Distanzen (B) der Daten aus Matrix 1. Je geringer der Abstand der Datenpunkte zweier Aminosäurepositionen ist, desto ähnlicher sind diese Positionen einander. Im gestrichelten Kreis befinden sich die beiden Kombinationen allgemeinvariabler Positionen und einige allgemeinvariable Einzelpositionen (siehe Text).

III.4.2 Zwei-Weg hierarchische Clusteranalyse von Matrix 2

III.4.2.1 Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher Rangkorrelation

Abbildung III.10 zeigt die Zwei-Weg hierarchische Clusteranalyse der Ähnlichkeitsmatrix, die auf Spearman'scher Rangkorrelation der Daten aus Matrix 2 beruht. Der Ähnlichkeitsbaum der Aminosäurepositionen dieser Abbildung ist in zwei Hauptäste unterteilt. Jeder Hauptast enthält einzigartige, charakteristische und allgemeinvariable Positionen. Es findet keine graphische Trennung der allgemeinvariablen, einzigartigen und charakteristischen Positionen statt. Insgesamt wurden acht Kombinationen von Positionen, die jeweils für ein Allel einzigartig sind, eine Kombination, die für zwei Allele charakteristisch ist und zwei Kombinationen, die allgemeinvariable Positionen enthalten, gebildet. Die Kombinationen sind in Tabelle III.7 dargestellt. In einigen Kombinationen einzigartiger Positionen weisen die Positionswerte starke Unterschiede auf. In Kombination 10 zum Beispiel befinden sich Aminosäurepositionen, die für das Allel B2 einzigartig sind. Dabei besitzt die Position 265 einen hohen Positionswert (98) und die Positionen 57 und 370 niedrige (10 und 30). In Tabelle III.6 sind die physikochemischen Differenzen aller Positionen dargestellt. Insgesamt nimmt der Kontrast der gelben Farbtöne im Vergleich zu der entsprechenden Analyse von Matrix 1 ab. Das bedeutet, dass bei vielen Aminosäuresubstitutionen nur geringe physikochemische Differenzen entstanden, so dass die entsprechenden Positionswerte klein sind.

Tabelle III.6: Aminosäurepositionen aus Matrix 2. Dargestellt sind die Merkmale der Aminosäurepositionen (einzigartig =E; charakteristisch =C; allgemeinvariabel =A) und die physikochemischen Differenzen. Weist eine Position mehrere physikochemische Differenzen auf, so werden diese durch ein Komma getrennt angegeben. Gegebenenfalls ist in Klammern dargestellt, für welche Allele die Positionen einzigartig bzw. charakteristisch sind. Enthält eine Position Deletionen, und sind an dieser Position mehr Aminosäuren deletiert als vorhanden, so wird das Vorhandensein angegeben. Sind mehr Aminosäuren vorhanden als deletiert, so werden die Deletionen angegeben.

Position	Merkmal	physikochemische Differenz	Position	Merkmal	physikochemische Differenz
7	A	46	307	Deletion bei B2	-
23	A	46	308	E (G3)	26
24	A	56	310	C (B3, R2, W2, Y1)	26
42	A	64	312	C (B2, R1, W1)	103, 74, 110
54	A	29	321	A	95
57	B2	10	322	A	92
83	R2	71	323	A	102, 94, 149
93	C (G1, G3, R2)	59	324	C (R1, R2)	144, 83, 89
104	Y3	23	352	E (W1)	95
144	Deletion bei R1	-	353	E (G3)	102
162	E (Y3)	32	356	E (B1)	85
173	C (Y1, Y3)	5	358	AS nur bei B1	-
190	A	57	359	AS nur bei B1, B3, R3, W1	29, 128
201	A	57	362	AS nur bei B1, B3, R3, W1	-
205	E (Y3)	95	368	C (W2, Y1, Y3)	30
211	Deletion bei W2	-	370	E (B2)	30
219	A	46	372	Deletion bei R2, W1, W2, Y1	5, 21, 22, 29, 32, 50
221	E (W2)	90	373	Deletion bei R2, W1, W2, Y1	-
230	C (G3, W2)	29, 98, 125	374	A	30
237	E (W1)	196	376	C (G3, R2, W2, Y1)	196
246	A	26	380	E (W2)	98
247	A	56	383	C (G1, Y2)	74
251	C (B1, W2)	43	385	C (G1, Y2)	98
256	E (W2)	175	387	E (G3)	40
259	E (W2)	57	388	C (G3, W1, Y2)	29, 21, 50
265	E (B2)	98	390	A	10
267	E (R2)	90	393	E (Y3)	29
269	E (G3)	181	399	A	83
270	C (G1, R2, Y2)	57	403	A	74
273	E (Y3)	156	405	E (G3)	74
281	C (B3, G2, R1, W2)	95	411	B1, W1, Y3	74
282	E (G3)	144	417	AS nur bei B1	-
287	A	86, 26, 94	418	E (B1)	21
288	A	80	419	C (B1, Y1)	29, 97
293	A	125	421	E (G3)	74
295	C (B3, R1, R3)	71	422	E (R1)	113
297	C (B2, R3)	26	427	C (B2, W2)	98
302	E (G3)	23	428	C (G2, G3)	98, 74, 145
303	E (R1)	26	429	E (G2)	59
304	C (B2, G2, R1, W1)	78, 71, 26	430	E (W1)	24
305	E (B1)	45	431	Deletion bei W1, B2	-

Einige Aminosäureunterschiede weisen so geringe physikochemische Differenzen auf, dass die entsprechenden Positionen dadurch bei bestimmten Allelen als nur noch gering verändert (Farbton bei einigen Allelen schwarzgelb, beim Rest der Allele nahezu schwarz, z.B. Position 388) bzw. fast gar nicht verändert (Farbton über alle Allele nahezu schwarz, z.B. Position 390) zu den anderen Allelen dargestellt werden. Eine Kombination allgemeinvariabler Positionen (Kombination 1 aus Tabelle III.7) enthält unter anderem die Positionen 358 und 417. Diese beiden Positionen weisen bei allen Allelen außer dem Allel B1 Deletionen auf, so dass aufgrund der Deletionen kein Vergleich zustande kam und damit keine physikochemischen Differenzen ermittelt werden konnten. Nicht bestimmbare physikochemische Differenzen an Positionen mit Deletionen sind in Abbildung III.10 an ihrer grauen Farbe erkennbar. Die acht Aminosäurepositionen, die nur bei einigen Allelen Deletionen aufweisen (144, 211, 307, 359, 362, 372, 373, und 431) (siehe Abschnitt III.5, Tabelle III.10), sind Einzelpositionen. Im Ähnlichkeitsbaum befinden sie sich in einem der beiden Hauptäste an den äußeren Verzweigungen mit großen Astlängen. Sie sind damit deutlich voneinander und von den anderen Positionen verschieden. Die physikochemischen Differenzen der Allele, die an diesen Positionen jeweils keine Deletionen aufweisen sind null. Die verbliebenen Allele einer Position haben daher alle die gleiche Rangzahl. Allele, die Deletionen aufweisen, gehen nicht in die Rangbildung ein. Das „Muster“ der Deletionen ist somit entscheidend für die Ähnlichkeit der Positionen, da unterschiedliche Allele in die Rangbildung eingehen. Position 372, die Deletionen bei den Allelen R2, W1, W2 und Y1 aufweist, ist der für das Allel W2 spezifischen Position 230 ähnlich. Position 359, die bei allen Allelen außer B1, B3, R3 und W1 Deletionen aufweist, ist der für die Allele B2 und R3 spezifischen Position 297 ähnlich. Die physikochemischen Differenzen der Allele, die an den Positionen 359 und 372 Positionen eine Aminosäure tragen, sind größer als null, da sich die Aminosäuren unterscheiden. Die Rangordnung der Allele den Positionen 359 und 372 ist nicht nur abhängig vom „Muster“ der Deletionen, sondern auch von den physikochemischen Differenzen, die die Allele an den beiden Position aufweisen. Die Aminosäurepositionen, die Deletionen aufweisen, werden im Anschluss an die ZWHC und MDS Analysen zusätzlich separat ausgewertet, um Informationen, die in den Deletionen enthalten sein könnten, zu bekommen.

Eine Übersicht über die Lokalisation der Aminosäuresubstitutionen und -deletionen der *csd*-Allele und die physikochemischen Differenzen der Aminosäuresubstitutionen ist in Abbildung III.15 gegeben.

Für die Allele B1, B2, G3, R1, R2, W1, W2 und Y3 gibt es Kombinationen einzigartiger Aminosäurepositionen (Tabelle III.7). Das Allel G2 hat eine einzigartige Einzelposition. Die Allele B3, G1, R3, Y1 und Y2 haben keine für sie einzigartigen Aminosäurepositionen. Bezüglich der Ähnlichkeit der Allele gibt es in Abbildung III.10. drei Cluster mit Allelen, die sich nach Berücksichtigung der physikochemischen Differenzen jeweils sehr ähnlich sind, nämlich die Allele R3 und Y1, die Allele B3 und R1 und die Allele G2 und Y2. Am ähnlichsten davon sind die Allele G2 und Y2, ihnen ist das Allel G1 ähnlich, das nur durch eine Verzweigung von G2 und Y2 getrennt ist. Am wenigsten Ähnlichkeit untereinander und mit den anderen Allelen weisen, nach Berücksichtigung der physikochemischen Differenzen, die Allele W2, G3, Y3 und B1 auf.

Tabelle III.7: Kombinationen von Aminosäurepositionen, die bei der ZWHC der Ähnlichkeitsmatrix, die auf Spearman'scher Rangkorrelation der Daten aus Matrix 2 beruht, gefunden wurden. Unter „Merkmal“ ist angegeben, ob die Kombination einzigartige (E), charakteristische (C) oder allgemeinvariable (A) Positionen enthält. Gegebenenfalls ist in Klammern angegeben, für welche Allele die Positionen einzigartig bzw. charakteristisch sind. Die physikochemischen Differenzen der Positionen werden in der gleichen Reihenfolge angegeben, wie die Positionen.

Nummer der Kombination	Positionen der Kombination	physikochemische Differenzen	Merkmal
1	7, 23, 24, 54	46, 46, 56, 29	A
2	303, 422	26, 113	E (R1)
3	269, 282, 302, 308, 353, 387, 405, 421	181, 144, 23, 26, 102, 40, 74, 74	E (G3)
4	383, 385	74, 98	C (G1, Y2)
5	237, 352, 430	196, 95, 24	C (W1)
6	246, 247	26, 56	A
7	83, 267	71, 90	E (R2)
8	104, 162, 205, 273, 393	23, 32, 95, 156, 29	E (Y3)
9	221, 256, 259, 380	90, 175, 57, 98	E (W2)
10	57, 265, 370	10, 98, 30	E (B2)
11	305, 356, 418	45, 85, 21	E (B1)

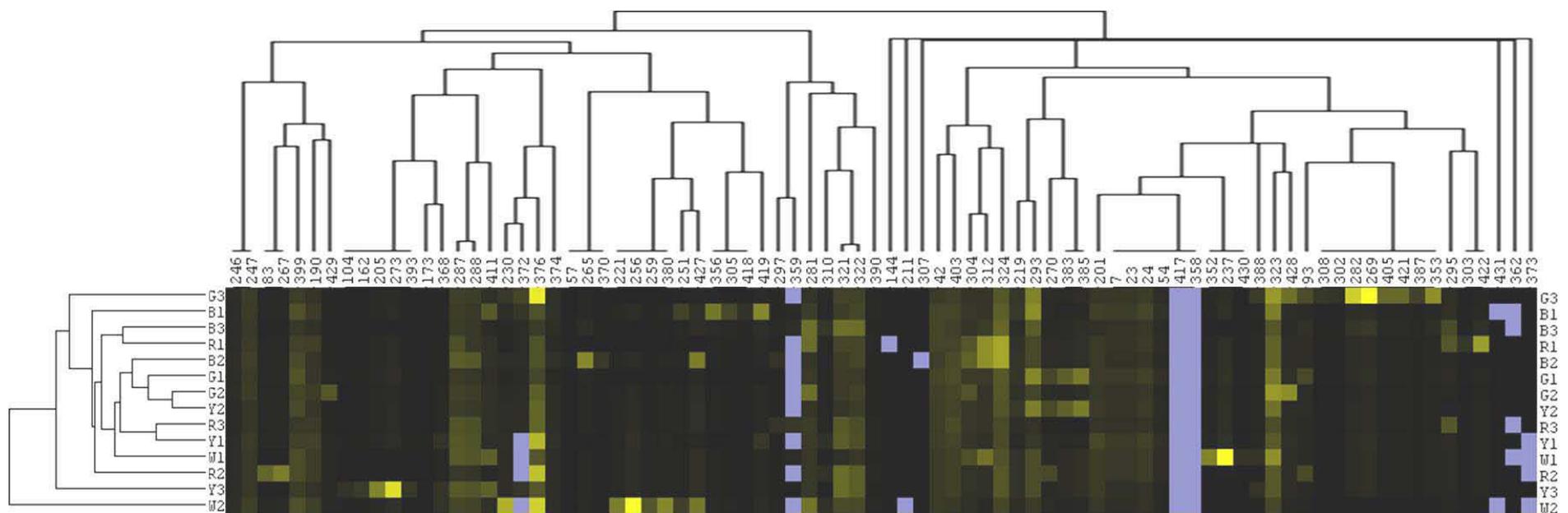


Abbildung III.10: Zwei-Weg Hierarchische Clusteranalyse von Matrix 2. Erstellung der Ähnlichkeitsmatrix erfolgte mittels Spearman'scher Rangkorrelation. Die Positionswerte aus Matrix 2 werden durch die Datenpunkte dargestellt. Jede Aminosäureposition in einem Allel entspricht einem Datenpunkt, dargestellt als Quadrat. Die Höhe der Positionswerte der Aminosäurepositionen wird durch die Farbintensität ausgedrückt. Je intensiver gelb das Quadrat, desto höher der Positionswert. Der Positionswert gibt den Wert der physikochemischen Differenz der betrachteten Position in dem betrachteten Allel an. Ist ein Datenpunkt ganz schwarz, so wurde das entsprechende Allel experimentell nicht mit dem Allel verglichen, welches an der betrachteten Aminosäureposition eine einzigartige Substitution aufweist. Bei Positionen, die Deletionen tragen können physikochemische Differenzen ebenfalls null betragen, wenn die nicht deletierten Aminosäuren an dieser Position sich nicht unterscheiden. Aufgrund von Deletionen fehlende Positionswerte sind grau. Der Ähnlichkeitsbaum der Allele und der der Aminosäurepositionen sind über die graphische Darstellung der Positionswerte verbunden, so dass ersichtlich ist, welche Aminosäurepositionen bei welchen Allelen hohe bzw. niedrige physikochemische Differenzen haben.

III.4.2.2 Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen

Abbildung III.11 zeigt die Clusteranalyse der Unähnlichkeitsmatrix, die auf Bildung der Euklidischen Distanzen der Daten aus Matrix 2 beruht. Der Ähnlichkeitsbaum der Aminosäurepositionen in Abbildung III.11 ist nicht in Hauptäste unterteilt. Aminosäurepositionen mit großen physikochemischen Differenzen werden durch frühe Verzweigung von Positionen mit sehr niedrigen physikochemischen Differenzen getrennt, unabhängig davon, für welche Allele sie einzigartig bzw. charakteristisch sind oder ob sie allgemeinvariabel sind. Allgemeinvariable und charakteristische Aminosäurepositionen werden in mehreren Clustern angeordnet. Die Aminosäurepositionen dieser Cluster weisen jeweils ähnliche physikochemische Differenzen auf. Die Positionen 54, 246, 374 und 310 befinden sich zum Beispiel in einem Cluster und weisen sehr geringe physikochemische Differenzen auf (26 bis 30). Die Positionen 7, 23, 24, 190, 201 und 247 eines anderen Clusters weisen physikochemische Differenzen in der Höhe von 46 bis 59 auf. Positionen mit niedrigen und gleichzeitig sehr ähnlichen physikochemischen Differenzen, die aber für unterschiedliche Allele einzigartig oder charakteristisch sind, werden entsprechend der Allele, für die sie einzigartig bzw. charakteristisch sind, in getrennte Cluster eingeordnet. Positionen, die für die gleichen Allele einzigartig oder charakteristisch sind und nur geringe Unterschiede in ihren physikochemischen Distanzen aufweisen, werden in Clustern angeordnet, in denen die terminalen Astlängen sehr kurz sind, so dass diese Positionen als sehr ähnlich dargestellt werden. Genauso werden Positionen dargestellt, die für unterschiedliche Allele einzigartig oder charakteristisch sind, deren physikochemische Differenzen aber sehr gering sind, so dass die Einzigartigkeit oder das Charakteristische der Positionen für die verschiedenen Allele kaum ausgeprägt ist. Diese Cluster mit sehr kurzen terminalen Astlängen, die sehr ähnliche Positionen enthalten, werden im Folgenden als Assoziationen ähnlicher Positionen bezeichnet. Die Assoziationen von Positionen sind in Tabelle III.8.A angegeben und nummeriert.

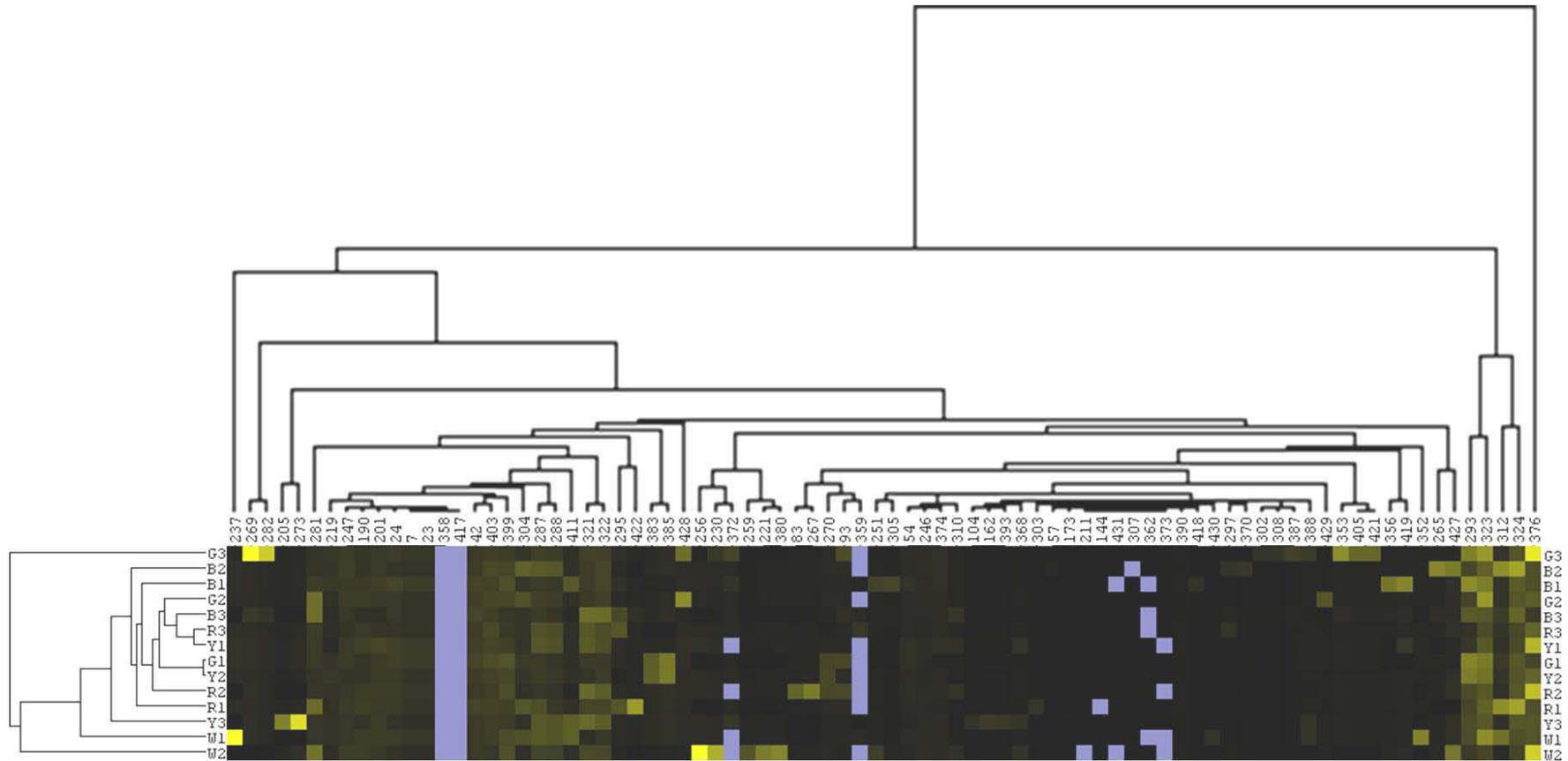


Abbildung III.11: Zwei-Weg Hierarchische Clusteranalyse von Matrix 2. Erstellung der Unähnlichkeitsmatrix erfolgte mittels Bildung der Euklidischen Distanzen. Die Positionswerte aus Matrix 2 werden durch die Datenpunkte dargestellt. Jede Aminosäureposition in einem Allel entspricht einem Datenpunkt, dargestellt als Quadrat. Die Höhe der Positionswerte der Aminosäurepositionen wird durch die Farbintensität ausgedrückt. Je intensiver gelb das Quadrat, desto höher der Positionswert. Der Positionswert gibt den Wert der physikochemischen Differenz der betrachteten Position in dem betrachteten Allel an. Ist ein Datenpunkt ganz schwarz, so wurde das entsprechende Allel experimentell nicht mit dem Allel verglichen, welches an der betrachteten Aminosäureposition eine einzigartige Substitution aufweist. Bei Positionen, die Deletionen tragen können physikochemische Differenzen ebenfalls null betragen, wenn die nicht deletierten Aminosäuren an dieser Position sich nicht unterscheiden. Positionswerte, die aufgrund von Deletionen fehlen, sind grau. Der Ähnlichkeitsbaum der Allele und der der Aminosäurepositionen sind über die graphische Darstellung der Positionswerte verbunden, so dass ersichtlich ist, welche Aminosäurepositionen bei welchen Allelen hohe bzw. niedrige physikochemische Differenzen haben.

Es werden bei der paarweisen Ähnlichkeitsdarstellung durch Bildung der Euklidischen Distanzen nur Positionen als genau gleich bezeichnet (einer Kombination zugeordnet), die bei den gleichen Allelen die gleichen physikochemischen Differenzen aufweisen. Das bedeutet, dass genau gleiche Aminosäurepositionen das gleiche Merkmal (einzigartig oder charakteristisch für die gleichen Allele oder gleiche allgemeine Variabilität) mit genau gleicher physikochemischer Differenz besitzen. In Abbildung III.11 ist zu erkennen, dass insgesamt nur drei Kombinationen von Aminosäurepositionen gebildet wurden. Kombination 1 enthält allgemeinvariable Positionen (7, 23) und die beiden Positionen 358 und 417, die aufgrund von Deletionen keine Positionswerte besitzen. Sechs der Aminosäurepositionen, die Deletionen aufweisen (144, 211, 307, 362, 373, und 431), werden Kombination 2 zugeordnet. Die Allele, die an diesen Positionen keine Deletion aufweisen, tragen die gleiche Aminosäure, so dass keine physikochemischen Differenzen auftreten. Die Aminosäurepositionen der Kombination 3 (405, 421) sind einzigartig für das Allel G3. Die beiden Positionen 372 und 359, die Deletionen bei den Allelen R2, W1, W2 und Y1 (Position 279) bzw. bei allen Allelen außer B1, B3, R3 und W1 (Position 359) aufweisen, werden keiner Kombination zugeordnet. Sie sind den Positionen 230 (charakteristisch für die Allele G3 und W2) bzw. 93 (charakteristisch für die Allele G1, G3 und R2) ähnlich. Die Aminosäuren der Allele, die an den Positionen 372 und 359 keine Deletion aufweisen, sind verschieden (siehe Alignment im Anhang Abbildung A.I), so dass zwischen diesen Aminosäuren physikochemische Differenzen bestehen. Diese physikochemischen Differenzen sind den entsprechenden physikochemischen Differenzen der Positionen 230 und 93 ähnlich.

Tabelle III.8 A: : Kombinationen von Positionen und Assoziationen von Positionen der ZWHC von Matrix 2 nach Bildung der Euklidischen Distanzen

Unter „Merkmal“ ist dargestellt, ob die Kombination oder die Assoziation einzigartige (E), charakteristische (C) oder allgemeinvariable (A) Positionen enthält. Gegebenenfalls ist in Klammern angegeben, für welche Allele die Positionen einzigartig bzw. charakteristisch sind. Falls sich die Merkmale von Positionen einer Assoziation voneinander unterscheiden, so werden die Merkmale in derselben Reihenfolge angegeben wie die Positionen, getrennt durch einen Schrägstrich. Die physikochemischen Differenzen sind ebenfalls in der gleichen Reihenfolge angegeben wie die Positionen. Positionen, die Deletionen aufweisen sind kursiv gedruckt.

Nummer der Kombination	Positionen der Kombination	physikochemische Differenzen	Merkmal
1	7, 23	46; 46	A
2	144, 211, 307, 362, 373, 431	0	-
3	405, 421	74; 74	E (G3)
Nummer der Assoziation	Positionen der Assoziation		
1	221, 380	90; 98	E (W2)
2	24, 190, 201, 247, Kombination 1	56; 57; 57; 56; 46	A
3	302, 308, 387	23; 26; 40	E (G3)
4	83, 267	71; 90	E (R2)
5	54, 246, 374	29; 26; 30	A
6	104, 162, 393	23; 32; 29	E (Y3)
7	57, 173, 390, 418, <i>Kombination 2</i>	10; 5; 10; 21; 0	E (B2)/A/A/ E (B1)/-

Aminosäurepositionen, die für die gleichen Allele einzigartig oder charakteristisch sind, aber deutliche Unterschiede in ihren physikochemischen Differenzen aufweisen, werden zwar in Clustern angeordnet, aber die terminalen Astlängen sind deutlich länger, als bei den Assoziationen von Positionen. Damit wurden die Aminosäurepositionen, die sich in diesen Clustern befinden als nicht so ähnlich erkannt, wie die Aminosäurepositionen, die sich in Assoziationen befinden. Diese Cluster, die Aminosäurepositionen enthalten, die für die gleichen Allele einzigartig oder charakteristisch sind bzw. die gleiche allgemeine Variabilität besitzen, aber deutlich unterschiedliche physikochemische Differenzen aufweisen, sind in Tabelle III.8.B angegeben.

Tabelle III.8 B: Positionen, die sich bei der ZWHC von Matrix 2 nach Bildung der Euklidischen Distanzen in Clustern mit langen terminalen Astlängen befinden.

Unter „Merkmal“ ist dargestellt, ob das Cluster einzigartige (E), charakteristische (C) oder allgemeinvariable (A) Positionen enthält. Gegebenenfalls ist in Klammern angegeben, für welche Allele die Positionen einzigartig bzw. charakteristisch sind. Falls sich die Merkmale von Positionen einer Assoziation voneinander unterscheiden, so werden Merkmale in derselben Reihenfolge angegeben wie die Positionen, getrennt durch einen Schrägstrich. Die physikochemischen Differenzen sind ebenfalls in der gleichen Reihenfolge angegeben wie die Positionen, getrennt durch ein Semikolon. Besitzt eine Position mehrere physikochemische Differenzen, so werden diese aufgezählt, getrennt durch Komma. Positionen, die Deletionen aufweisen sind kursiv gedruckt.

Nummer des Clusters	Positionen des Clusters	physikochemische Differenz	Merkmal
1	256, 230, 372	175; 29, 98, 125; 5, 21, 22, 29, 32, 50	E (W2)/E (G3), E (W2) /-
2	259, Assoziation 1	57; 90, 98	E (W2)
3	205, 273	95; 156	E (Y3)
4	353, Kombination 3	102; 74	E (G3)
5	383, 385	74; 98	C (G1, Y2)
6	321, 322	95; 92	A
7	287, 288	85, 26, 94; 80	A
8	42, 403	64; 74	A
9	269, 282	181, 144	E (G3)
10	219, Assoziation 2	46; 46, 56, 57	A
11	297, 370	26; 30	C (B2, R3)/E (B2)
12	93, 270, 359	59; 57; 29, 128	C (G1, G3, R2)/ C (G1, R2, Y2)/-

Die Eigenschaften der Aminosäurepositionen, die nach der ZWHC der Daten aus Matrix 2 (nach Bildung der Euklidischen Distanzen) Kombinationen, Assoziationen und Clustern mit langen terminalen Astlängen zugeordnet wurden, sind zur Veranschaulichung in Abbildung III.12. dargestellt.

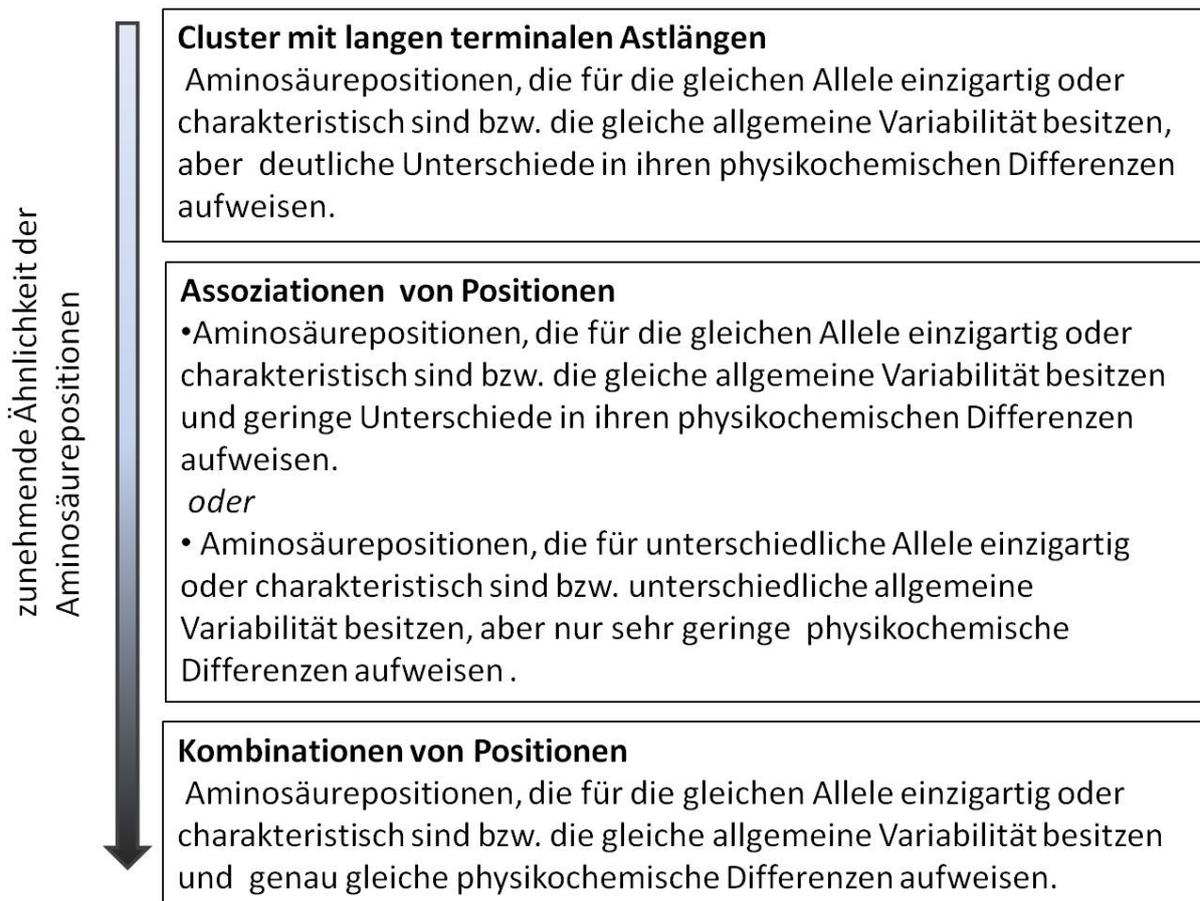


Abbildung III.12: Einordnung von Positionen in Cluster, Assoziationen und Kombinationen bei der ZWHC nach Bildung der Euklidischen Distanzen der Daten aus Matrix 2.

Wie oben erwähnt, werden Aminosäurepositionen mit großen physikochemischen Differenzen im Baum früh von Positionen mit niedrigen physikochemischen Differenzen getrennt, unabhängig davon, für welche Allele sie einzigartig oder charakteristisch sind. Aminosäurepositionen die für die gleichen Allele einzigartig oder charakteristisch sind, deren physikochemische Differenzen sich in der Höhe aber sehr stark unterscheiden, werden daher im Baum nicht als ähnlich dargestellt. Dies ist z.B. der Fall bei den Positionen 303 und 422. Sie sind beide spezifisch für das Allel R1. Position 303 hat bei R1 eine relativ niedrige physikochemische Differenz (26), Position 422 hat eine hohe (113). Bei manchen Positionen sind die physikochemischen Differenzen so gering, dass geringvariable einzigartige oder charakteristische Positionen mit geringvariablen allgemeinvariablen Positionen in eine Assoziation eingeordnet werden. Dies ist z.B. der Fall bei den Positionen aus Assoziation 7 (Tabelle III.8a). Der Wert der physikochemischen Differenzen ist offensichtlich die entscheidende Kenngröße, die bei der Bildung der Euklidischen Distanzen die Ähnlichkeit der

Aminosäurepositionen bestimmt. Die Merkmale der Positionen (einzigartig, charakteristisch oder allgemeinvariabel) werden erst an zweiter Stelle berücksichtigt.

Wie in Tabelle III.8.A und B und Tabelle III.6 zu sehen ist, haben die Allele B1, B2, G2, G3, R1, R2, W1, W2 und Y3 nach Berücksichtigung der physikochemischen Differenzen einzigartige Positionen in Form von Kombinationen, Assoziationen, und / oder einer Einzelposition. Die Allele B3, G1, R3, Y1 und Y2 besitzen keine einzigartigen Aminosäurepositionen. In Abbildung III.11 ist zu erkennen, dass die Allele R3, Y1 und B3 sowie die Allele G1 und Y2 im Ähnlichkeitsbaum der Allele jeweils Cluster bilden. Die Allele G1 und Y2 sind sich nach Berücksichtigung der physikochemischen Differenzen so ähnlich, dass die terminalen Astlängen fast null betragen. Die Allele G3, W2, W1 und Y3 haben nach Berücksichtigung der physikochemischen Differenzen die geringste Ähnlichkeit zu den jeweils 13 anderen Allelen.

III.4.3 Analysen der Matrix 3

III.4.3.1 Multidimensionale Skalierung der Allele von Matrix 3

III.4.3.1.1 Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher Rangkorrelation

In Abbildung III.13.A ist die Multidimensionale Skalierung der Allele gezeigt, die auf Spearman'scher Rangkorrelation der Daten aus Matrix 3 beruht. Der Stresswert beträgt 0,17. Die Ordination ist daher als nicht sehr zuverlässig einzustufen. Es kann nicht davon ausgegangen werden, dass die Ähnlichkeiten der Allele detailgetreu wiedergegeben werden. Die Datenpunkte, die die Allele darstellen sind nahezu kreisförmig über die Fläche der Graphik verteilt. Die Datenpunkte der Allele G3, W2 und Y3 liegen außerhalb des Kreises in den Randgebieten der Graphik. Diese drei Allele sind somit den anderen Allelen unähnlich. Die Rangordnungen der Positionen nach Berücksichtigung der physikochemischen Differenzen sind sich bei diesen Allelen offensichtlich unähnlich. Am ähnlichsten zueinander sind die Allele G2 und Y1 und Y2, denn ihre Datenpunkte haben die geringsten Abstände zueinander. Die Rangordnungen der Positionen dieser Allele sind sich also nach Berücksichtigung der physikochemischen Differenzen am ähnlichsten. Die genauen Koordinatenwerte der Datenpunkte der Allele sind in Tabelle A.XII angegeben.

III.4.3.1.2 Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen

In Abbildung III.13.B ist die Multidimensionale Skalierung der Allele gezeigt, die auf Bildung der Euklidischen Distanzen der Daten aus Matrix 3 beruht. Mit einem Stresswert von 0,08 ist diese Ordination als zuverlässig einzustufen. Die Datenpunkte der Allele B3, G1, G2, R3, Y1 und Y2 liegen im Zentrum der Graphik sehr eng zusammen. Diese Allele sind sich also sehr ähnlich. Das bedeutet, dass diese Allele wenige einzigartige Positionen aufweisen und / oder einzigartige Positionen nur geringe physikochemische Differenzen besitzen und die Euklidischen Differenzen daher klein bleiben. Kreisförmig um dieses Zentrum herum liegen die Datenpunkte der Allele B1, B2, R1, R2 und Y3. Am wenigsten Ähnlichkeit zu den jeweils anderen 13 Allelen haben die Allele G3, W1 und W2, denn ihre Datenpunkte liegen in den Randgebieten der Graphik. Diese Allele weisen viele einzigartige Positionen auf, an denen

die physikochemischen Differenzen einen hohen Wert haben, was zu großen Euklidischen Distanzen führt. Die genauen Koordinatenwerte der Datenpunkte der Allele sind in Tabelle A.XIII angegeben.

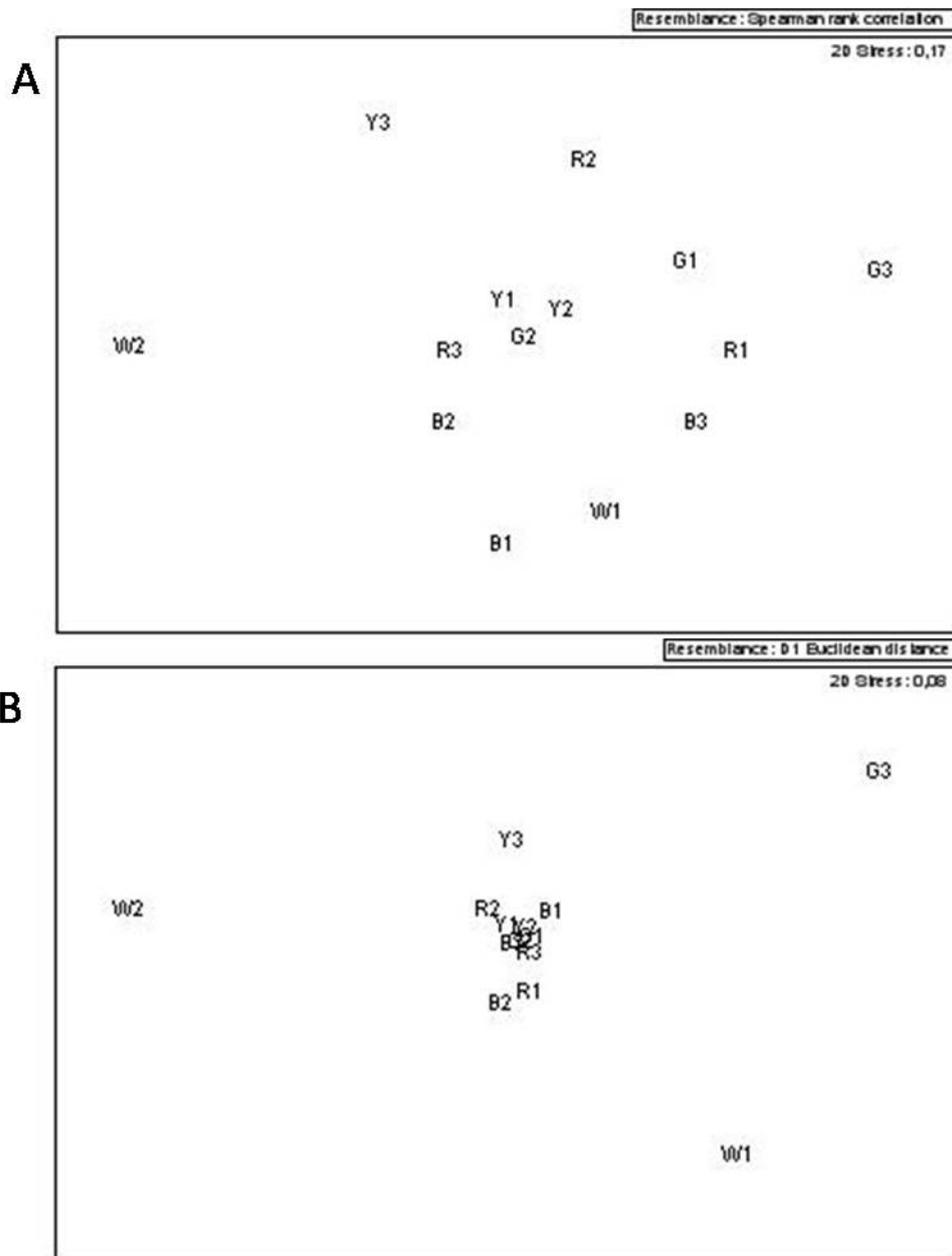


Abbildung III.13: Multidimensionale Skalierung der Allele aus Matrix 3. Berechnung der (Un)Ähnlichkeitsmatrix mittels Spearman'scher Rangkorrelation (A) und Bildung der Euklidischen Distanzen (B) der Daten aus Matrix 3. Je geringer der Abstand der Datenpunkte zweier Allele ist, desto ähnlicher sind diese Allele einander.

III.4.3.2 Multidimensionale Skalierung der Aminosäurepositionen von Matrix 3

III.4.3.2.1 Analyse der Ähnlichkeitsmatrix basierend auf Spearman'scher Rangkorrelation

In Abbildung III.14.A ist die Multidimensionale Skalierung der Aminosäurepositionen gezeigt, die auf Spearman'scher Rangkorrelation der Daten aus Matrix 3 beruht. Mit einem Stresswert von 0,25 kann nicht davon ausgegangen werden, dass die Ordination zuverlässige Angaben zur Ähnlichkeit der Positionen zeigt. Die Graphik wird daher nicht in die Analyse der Ähnlichkeit der Aminosäurepositionen einbezogen.

III.4.3.2.2 Analyse der Unähnlichkeitsmatrix basierend auf Bildung der Euklidischen Distanzen

In Abbildung III.14.B ist die Multidimensionale Skalierung der Aminosäurepositionen gezeigt, die auf Bildung der Euklidischen Distanzen der Daten aus Matrix 3 beruht. Die Ordination hat einen Stresswert von 0,15 und kann somit herangezogen werden, um die Größenordnungen der Ähnlichkeiten der Aminosäurepositionen zu erkennen.

Es wurden drei Kombinationen von Aminosäurepositionen gebildet, deren Datenpunkte identische Koordinatenwerte aufweisen (Tabelle III.9). Aminosäurepositionen, deren Datenpunkte identische Koordinatenwerte aufweisen, haben bei jedem Allel genau gleich hohe Positionswerte. Das bedeutet, dass diese Aminosäurepositionen bei allen Allelen identische physikochemische Differenzen aufweisen. Außerdem lassen sich zehn Kombinationen mit Aminosäurepositionen erkennen, deren Datenpunkte zwar sehr ähnliche, aber nicht identische Koordinatenwerte aufweisen. Diese Positionen sind sich ähnlich. Sie haben entweder bei den gleichen Allelen ähnliche physikochemische Differenzen oder ihre physikochemischen Differenzen sind bei allen Allelen so gering, dass aus diesen Gründen ihre Euklidische Distanz nur klein ist. In der ZWHC von Matrix 2 nach Bildung der Euklidischen Distanzen wurden solche Positionen in Clustern angeordnet, deren terminale Astlängen nur sehr klein waren. Die Positionen dieser Cluster wurden als Assoziation von ähnlichen Positionen bezeichnet. Dieser Begriff wird nun auch für Positionen genutzt, deren Datenpunkte bei der MDS sehr ähnliche Koordinatenwerte haben. Die Einteilung von Positionen in Assoziationen nach Multidimensionaler Skalierung ist nur eine Grobeinteilung.

Die Position 418 könnte anhand der Ähnlichkeit ihrer Koordinatenwerte zwei verschiedenen Assoziationen zugeordnet werden. Der Datenpunkt von Position 418 ist sowohl den Datenpunkten der Positionen 57 und 173 nahe als auch dem der Position 370. Die von mir in Tabelle III.9 festgehaltene Einteilung in Assoziationen ist somit nicht die einzig mögliche.

Die Ähnlichkeit der Positionen zueinander wird bei dieser Analyse vor allem durch die Ähnlichkeit der physikochemischen Differenzen, die die Positionen besitzen, bestimmt. Erst bei ähnlichen Werten der physikochemischen Differenzen der Positionen wird ermittelt, ob die Positionen gleiche Merkmale (einzigartig bzw. charakteristisch für (ein) bestimmte(s) Allel(e) oder gleiche allgemeine Variabilität) aufweisen. Die Positionen 246, 297, 303 und 308 haben zum Beispiel ähnliche physikochemische Differenzen (26), sind aber für unterschiedliche Allele einzigartig, charakteristisch oder allgemeinvariabel. Ihre Datenpunkte liegen deswegen nicht so nahe zusammen, dass sie einer Assoziation zugeordnet werden können. Sie werden daher in unterschiedliche Assoziationen eingeordnet. Die Datenpunkte der Kombinationen und Assoziationen von Positionen wurden graphisch in einer „Wolke“ angeordnet. Diese befindet sich an einer Seite der Graphik. Die Datenpunkte der anderen Aminosäurepositionen wurden in unterschiedlichen Abständen zu dieser Wolke angeordnet. Die Datenpunkte der Aminosäurepositionen 237, 256, 269 und 376 befinden sich am weitesten von der Wolke entfernt in den Randgebieten der Graphik. Diese Positionen haben große physikochemische Differenzen und sind daher den anderen Aminosäurepositionen sehr unähnlich. Alle Kombinationen und Assoziationen von Positionen sind in Tabelle III.9 angegeben. Die Koordinatenwerte der Datenpunkte aller Positionen sind in Tabelle A.XIV angegeben.

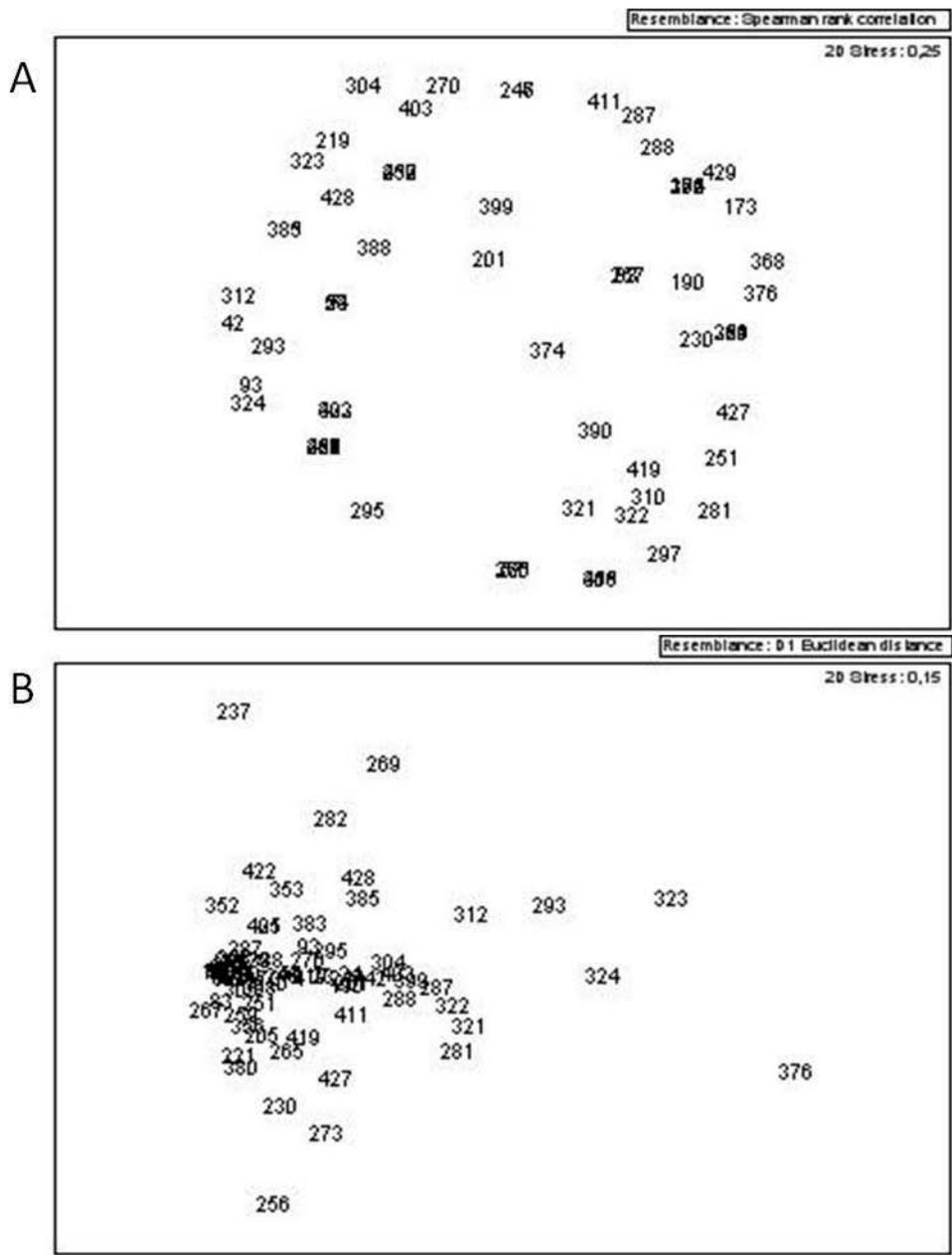


Abbildung III.14: Multidimensionale Skalierung der Aminosäurepositionen aus Matrix 3. Berechnung der (Un)Ähnlichkeitsmatrix mittels Spearman'scher Rangkorrelation (A) und Bildung der Euklidischen Distanzen (B) der Daten aus Matrix 3, in der alle Aminosäurepositionen, die Deletionen aufweisen, entfernt wurden. Je geringer der Abstand zweier Datenpunkte ist, desto ähnlicher sind die entsprechenden Aminosäurepositionen einander.

Tabelle III.9: Multidimensionale Skalierung der Aminosäurepositionen: Die Unähnlichkeitsmatrix wurde mittels Bildung der Euklidischen Distanzen der Daten von Matrix 3 berechnet. Dargestellt sind Kombinationen und Assoziationen von Aminosäurepositionen. In Kombinationen befinden sich Aminosäurepositionen mit identischen Koordinatenwerten. Assoziationen wurden Aminosäurepositionen mit sehr ähnlichen Koordinatenwerten zugeordnet. Unter „Merkmal“ ist angegeben, ob die Kombination oder die Assoziation einzigartige (E), charakteristische (C) oder allgemeinvariable (A) Positionen enthält. Gegebenenfalls ist in Klammern angegeben, für welche Allele die Positionen einzigartig bzw. charakteristisch sind. Gegebenenfalls ist in Klammern angegeben, für welche Allele die Positionen einzigartig bzw. charakteristisch sind. Falls sich die Merkmale von Positionen einer Assoziation voneinander unterscheiden, so werden die Merkmale in derselben Reihenfolge angegeben wie die Positionen, getrennt durch einen Schrägstrich. Die physikochemischen Differenzen sind ebenfalls in der gleichen Reihenfolge angegeben wie die Positionen, getrennt durch ein Semikolon. Besitzt eine Position mehrere physikochemische Differenzen, so werden diese aufgezählt, getrennt durch Komma.

Nummer der Kombination	Positionen der Kombination	physikochemische Differenz	Merkmal
1	7, 23	46; 46	A
2	405, 421	74; 74	E (G3)
3	393, 162	32; 29	E (Y3)
Nummer der Assoziation	Positionen der Assoziation		
1	Kombination 3, 104	32, 29; 23	E (Y3)
2	390, 297	10, 26	A/C (B2, R3)
3	303, 302, 308, 430	26; 23; 26; 24	E (R1)/E (G3)/E (G3)/E (W1)
4	54, 246, 374	29; 26; 30	A
5	57, 173	10; 5	E (B2)/C (Y1, Y3)/E (B2)
6	370, 418	30; 21	E (B2)/E (B1)
7	24, 190, 201, 247	56; 57; 56	A
8	221, 380	90; 98	E (W2)
9	251, 388	43; 29, 21, 50	C (B1, W2)/C (G3, W1, Y2)
10	93, 270	59; 57	C (G1, G3, R2)/C (G1, R2, Y2)

III.5 Separate Analyse der Aminosäurepositionen, die Deletionen aufweisen

Die Aminosäurepositionen 144, 211, 307, 358, 359, 362, 372, 373, 417 und 431 weisen Deletionen bei einem bis 13 Allelen auf (Tabelle III.10). Die Eigenschaften dieser Aminosäurepositionen werden nun in einer separaten Auswertung dargestellt. Als einzigartig oder charakteristisch für (ein) Alle(e) wird das Fehlen oder das Vorhandensein von Aminosäuren bezeichnet. Sind an einer Position mehr Aminosäuren deletiert, als vorhanden sind, so wird das Vorhandensein der Aminosäure als einzigartiges oder charakteristisches Merkmal bezeichnet. Sind an einer Position mehr Aminosäuren vorhanden als deletiert sind, so werden die Deletionen als einzigartiges oder charakteristisches Merkmal angegeben. Weist zum Beispiel eine Aminosäureposition bei zwei der 14 Allele eine Deletion auf, so werden diese Deletionen als charakteristisch für die beiden Allele bezeichnet. Die Deletionen bzw. die Aminosäuren, die die Positionen für bestimmte Allele einzigartig oder charakteristisch machen, sind in Tabelle III.10 farbig hervorgehoben.

Tabelle III.10: Merkmale der Aminosäurepositionen, die Deletionen tragen.

Weist eine Aminosäureposition Deletion(en) bei einem oder mehreren Allelen auf, so ist dies mit ~ gekennzeichnet. Die einzigartigen oder charakteristischen Merkmale (in diesem Fall das Vorhandensein oder die Deletion von Aminosäuren) der Aminosäurepositionen sind rot gekennzeichnet. Sind an einer Position mehr Aminosäuren deletiert, als vorhanden sind, so wird das Vorhandensein der Aminosäure als einzigartiges oder charakteristisches Merkmal bezeichnet. Sind an einer Position mehr Aminosäuren vorhanden als deletiert sind, so werden die Deletionen als einzigartiges oder charakteristisches Merkmal angegeben.

A		POSITIONEN									
		144	211	307	358	359	362	372	373	417	431
ALLELE	B1	T	A	E	Y	I	~	I	P	Y	~
	B2	T	A	~	~	~	Y	V	P	~	Q
	B3	T	A	E	~	I	~	V	P	~	Q
	G1	T	A	E	~	~	Y	V	P	~	Q
	G2	T	A	E	~	~	Y	V	P	~	Q
	G3	T	A	E	~	~	Y	F	P	~	Q
	R1	~	A	E	~	~	Y	V	P	~	Q
	R2	T	A	E	~	~	Y	~	~	~	Q
	R3	T	A	E	~	V	~	V	P	~	Q
	W1	T	A	E	~	I	~	~	~	~	Q
	W2	T	~	E	~	~	Y	~	~	~	~
	Y1	T	A	E	~	~	Y	~	~	~	Q
	Y2	T	A	E	~	~	Y	V	P	~	Q
	Y3	T	A	E	~	~	Y	L	P	~	Q

Wie in Tabelle III.10. ersichtlich ist, haben die Aminosäurepositionen 144, 211 und 307 bei jeweils einem Allel Deletionen, die Position 431 bei zwei Allelen. Die Positionen 358 und 417 tragen ausschließlich bei dem Allel B1 Aminosäuren, die anderen Allele weisen an diesen Positionen eine Deletionen auf. An Position 359 befinden sich nur bei den Allelen B1, B3, R3 und W1 Aminosäuren. Das Allel R3 weist an dieser Position ein Valin auf, die Allele B1, B3 und W1 ein Isoleucin. Die physikochemische Differenz zwischen Valin und Isoleucin ist mit 29 gering. Allele, deren Position 362 eine Aminosäure trägt, weisen an Position 359 eine Deletion auf. Die Allele, die an Position 359 keine Deletion aufweisen, tragen stattdessen ein Tyrosin an dieser Stelle. An den Positionen 372 und 373 tragen die Allele R2, W1, W2 und Y1 keine Aminosäure. Die Allele, die an diesen beiden Positionen keine Deletion aufweisen, tragen an Position 373 ein Prolin und an Position 372 eine von vier verschiedenen Aminosäuren. Bei Allel B1 ist dies ein Isoleucin, bei den Allelen B2, B3, G1, G2, R1, R3 und Y2 ein Valin, bei Allel G3 ein Phenylalanin und bei dem Allel Y3 ein Leucin. Die physikochemischen Differenzen der Aminosäuren reichen von 5 (Leucin - Isoleucin) bis 50 (Valin - Phenylalanin), sind also verhältnismäßig gering.

Diese Auswertung zeigt, dass die von Deletionen betroffenen Positionen für bestimmte Allele einzigartig oder charakteristisch sind. Die Deletionen könnten neben den Aminosäuresubstitutionen, die in den vorigen Analysen beschrieben wurden, eine Rolle bei der Entstehung der Spezifität von *csd*-Allelen spielen.

III.6 Vergleich der Ergebnisse der Zwei-Weg hierarchischen Clusteranalyse und der Multidimensionalen Skalierung

Die Analyseergebnisse der drei Matrices werden verglichen, um festzustellen, inwiefern sich die gefundenen Ergebnisse unterscheiden bzw. gleichen.

III.6.1 Vergleich der Ergebnisse der Allelanalysen

Die Ergebnisse zeigen, dass die Darstellungen der Ähnlichkeiten bzw. Unähnlichkeiten der Allele mittels Zwei-Weg hierarchischer Clusteranalyse und mittels Multidimensionaler Skalierung sich bei sonst konstanten Analyseparametern (Matrix und (Un)ähnlichkeitsmatrix) wenig unterscheiden.

Die Darstellung der Ähnlichkeit der Allele wird bei den Analysen der drei verschiedenen Matrices 1, 2 und 3 in Abhängigkeit von der Wahl der Ähnlichkeitsbestimmung (Spearman'sche Rangkorrelation oder Bildung der Euklidischen Distanzen) unterschiedlich dargestellt. Bei den Zwei-Weg hierarchischen Clusteranalysen der Matrix 1 sind die Allele G1, Y2 und G2 sehr ähnlich, sowohl nach Bildung der Euklidischen Distanzen (Abbildung III.7) als auch nach Spearman'scher Rangkorrelation (Abbildung III.6). Die Allele W2, G3, B1 und Y3 sind untereinander und zu den übrigen Allelen am wenigsten ähnlich. Nach Berücksichtigung der physikochemischen Differenzen sind bei den Zwei-Weg hierarchischen Clusteranalysen der Matrix 2 nach Spearman'scher Rangkorrelation (Abbildung III.10) die Allele G2 und Y2 und nach Bildung der Euklidischen Distanzen (Abbildung III.11) die Allele G1 und Y2 am ähnlichsten. Die Allele W2 und G3 haben die geringste Ähnlichkeit zu den anderen Allelen. Bei der Multidimensionalen Skalierung der Matrix 1 nach Spearman'scher Rangkorrelation (Abbildung III.8.A) sind die Datenpunkte der Allele B2, B3, G1, G2, R1, R2, R3, Y1 und Y2 fast homogen um das Zentrum der Graphik verteilt mit meist gleichen Abständen zueinander. Die Allele B1, G3, W1, W2 und Y3 liegen etwas weiter in den Randgebieten und sind somit untereinander und zu den anderen Allelen am wenigsten ähnlich. Bei der Multidimensionalen Skalierung von Matrix 3 nach Spearman'scher Rangkorrelation (Abbildung III.13.A) ist zu erkennen, dass die Datenpunkte der Allele B2, G2, R3, Y1 und Y2 im Zentrum der Graphik etwas enger zusammen liegen als bei Matrix 1. Da die Stresswerte für die Ordinationen 0,18 bzw. 0,17 betragen, kann nicht davon ausgegangen werden, dass die Ähnlichkeiten der Allele in den Ordinationen bis ins Detail genau wiedergegeben werden. Die Unterschiede in den Ähnlichkeitsdarstellungen mittels könnten daher durch die Ungenauigkeit der Darstellungsweise entstanden sein. Die Allele B1, G3, R2, W1, W2 und Y3 liegen in beiden Ordinationen in den Randgebieten. Ein deutlicher Unterschied zwischen der MDS von Matrix 1 und Matrix 3 nach Spearman'scher Rangkorrelation ist die Ähnlichkeitsdarstellung der Allele B1 und W1. Bei der MDS von Matrix 3 haben die Allele B1 und W1 einen viel geringeren Abstand zueinander als bei der MDS von Matrix 1. Das bedeutet, dass diese Allele durch Berücksichtigung der physikochemischen Differenzen und / oder das Fehlen der Positionen, die Deletionen tragen, in Matrix 3 den anderen Allelen ähnlicher sind, als bei alleiniger Betrachtung der Anzahl der Aminosäureunterschiede in Matrix 1. Bei der MDS der Allele von Matrix 1 nach Bildung der Euklidischen Distanzen (Abbildung III.8.B) sind Allele, die sich untereinander ähneln, deutlich im Zentrum

zusammengezogen und stärker von Allelen separiert, die zu allen anderen Allelen unähnlich sind (Allele in den Randgebieten der Graphik), als nach Spearman'scher Rangkorrelation (Abbildung III.8A). Bei der MDS von Matrix 3 nach Bildung der Euklidischen Distanzen kommt dies noch stärker zum Ausdruck. Ähnliche Allele werden im Zentrum stark zusammengeballt und Allele, die zu den jeweils anderen 13 Allelen wenig Ähnlichkeit aufweisen, liegen deutlich in Randgebieten der Graphik. Die Berücksichtigung der physikochemischen Differenzen führt offensichtlich dazu, dass Allele als ähnlicher oder unähnlicher dargestellt werden, als bei der Analyse der Aminosäuresubstitutionen ohne Berücksichtigung der physikochemischen Differenzen in Matrix 1.

Bei der MDS von Matrix 1 nach Bildung der Euklidischen Distanzen werden die Datenpunkte der Allele B3, G1, G2, R2, R3, Y1 und Y2 sehr eng zusammenliegend angeordnet. Die Datenpunkte der Allele B1, B2, G3, W1, W2 und Y3 liegen in den Randgebieten. Nach Berücksichtigung der physikochemischen Differenzen sind die Allele B1, B2 und Y3 bei der ZWHC von Matrix 2 (Abbildung III.11) und der MDS von Matrix 3 (Abbildung III.13) nach Bildung der Euklidischen Distanzen den übrigen Allelen deutlich ähnlicher als bei den entsprechenden Analysen von Matrix 1 und den Analysen von Matrix 2 und Matrix 3 nach Spearman'scher Rangkorrelation (Abbildungen III. 10 und III.13.A). Das bedeutet, dass die Berücksichtigung der physikochemischen Differenzen dazu führt, dass die Allele B1, B2 und Y3 den anderen Allelen ähnlicher sind, als nur bei Betrachtung der Anzahl der Aminosäureunterschiede oder der Ränge, die die Positionen bei den verschiedenen Allelen haben.

Eine Einteilung der *csd*-Allele in Typen war basierend auf den Ergebnissen der ZWHC und der MDS der Matrices 1, 2 und 3 nicht möglich.

III.6.2 Vergleich der Ergebnisse der Analysen der Aminosäurepositionen

Wie sehr die Aminosäurepositionen für bestimmte Allele einzigartig oder charakteristisch sind, wurde bei Matrix 2 und Matrix 3, im Gegensatz zu Matrix 1, durch Berücksichtigung der physikochemischen Differenzen quantifiziert. Dies hatte zur Folge, dass bei Matrix 2 und Matrix 3 bei einigen Aminosäurepositionen kaum noch erkennbar war, dass diese einzigartig oder charakteristisch sind, da die substituierten Aminosäuren nur geringe physikochemische Differenzen zu den Aminosäuren der anderen Allele an dieser Position aufwiesen. Bei

anderen Aminosäurepositionen wurde dagegen aufgrund großer physikochemischer Differenzen stark betont, dass diese einzigartig oder charakteristisch sind.

Ein Vergleich der Ähnlichkeitsdarstellungen der Aminosäurepositionen, die bei den Zwei-Weg Hierarchischen Clusteranalysen von Matrix 1 und Matrix 2 nach Berechnung der Ähnlichkeitsmatrix mittels Spearman'scher Rangkorrelation gebildet wurden (Abbildungen III.6 und III.10), zeigt nur eine geringe Änderung der Ähnlichkeit einiger Aminosäurepositionen zueinander. Die Einteilung der Aminosäurepositionen in Kombinationen erfolgt bei den Analysen der Matrices 1, 2 und 3 nach Spearman'scher Rangkorrelation danach, für welche Allele die Positionen einzigartig oder charakteristisch sind oder ob sie die gleiche allgemeine Variabilität aufweisen, unabhängig davon, wie groß die physikochemische Differenz ist. Die Kombinationen von Aminosäurepositionen sind daher bei den Matrices 1, 2 und 3 identisch (Tabelle III.11).

Die Darstellung der Ähnlichkeit der Aminosäurepositionen zueinander nach Bildung der Euklidischen Distanzen zeigt starke Unterschiede zwischen den Analysen von Matrix 1 (ZWHC, Abbildung III.7 und MDS, Abbildung III.9.B) und den Analysen von Matrix 2 (ZWHC, Abbildung III.11) und Matrix 3 (MDS, Abbildung III.14.B).

Nach Bildung der Euklidischen Distanzen werden bei den Analysen von Matrix 1 die allgemeinvariablen Aminosäurepositionen größtenteils von den einzigartigen und charakteristischen Aminosäurepositionen graphisch separiert, insbesondere von den Kombinationen einzigartiger und charakteristischer Aminosäurepositionen. Die große Spannweite der physikochemischen Differenzen in Matrix 2 und Matrix 3 wirkt sich direkt auf die Spannweite der Euklidischen Distanzen aus. Die großen Unterschiede bei den physikochemischen Differenzen führen bei der MDS von Matrix 3 und der ZWHC von Matrix 2 dazu, dass die Ähnlichkeit der Positionen nach Bildung der Euklidischen Distanzen primär nach den Werten ihrer physikochemischen Differenzen beurteilt wird, das Merkmal der Positionen ist zweitrangig. Dies führt dazu, dass die Euklidischen Distanzen bei Positionen, die für die gleichen Allele einzigartig oder charakteristisch sind, aber unterschiedliche physikochemische Differenzen haben, hohe Werte aufweisen können. Positionen mit hohen physikochemischen Differenzen werden als unähnlich zu den anderen Positionen dargestellt, falls diese nicht bei den gleichen Allelen ebenso hohe physikochemische Differenzen aufweisen. Die Datenpunkte der Positionen 237 (W1), 256 (W2), 269 (G3) und 376 (G3, R2, W2 und Y1) mit ihren hohen physikochemischen Differenzen (196, 175, 181 und 196) liegen

z.B. in der MDS von Matrix 3 in den Randgebieten der Graphik. Diese Positionen sind somit den anderen Positionen unähnlich.

Sowohl bei der ZWHC von Matrix 2 (Tabelle III.8) als auch bei der MDS von Matrix 3 (Tabelle III.9) nach Bildung der Euklidischen Distanzen wurden einige Aminosäurepositionen, die sich bei allen Analysen von Matrix 1 in getrennten Kombinationen befanden oder Einzelpositionen waren, gemeinsam in Assoziationen von Positionen (ZWHC und MDS) oder Cluster mit langen terminalen Astlängen (ZWHC) eingeordnet. Assoziationen, die Positionen enthalten, die sich bei den Analysen von Matrix 1 in unterschiedlichen Kombinationen befanden oder Einzelpositionen waren, werden als neugeordnete Assoziationen oder Cluster bezeichnet und sind in Tabelle III.11 als „Neuordnungen“ angegeben. Es ist zu erkennen, dass die Neuordnungen, die bei der ZWHC von Matrix 2 nach Bildung der Euklidischen Distanzen entstanden, sich größtenteils von den Neuordnungen unterscheiden, die bei der MDS von Matrix 3 nach Bildung der Euklidischen Distanzen erzeugt wurden. Während sich in den Neuordnungen der ZWHC von Matrix 2 meist allgemeinvariable Positionen befinden, bestehen die Neuordnungen der MDS von Matrix 3 hauptsächlich aus einzigartigen und charakteristischen Positionen, die ganz oder teilweise für unterschiedliche Allele einzigartig oder charakteristisch sind. Dies zeigt, dass die Darstellungen der Ähnlichkeiten von Positionen durch die ZWHC und durch die MDS deutlich voneinander abweichen können, wenn die Ähnlichkeiten nicht eindeutig sind. Eindeutig sind Ähnlichkeiten nur, wenn Positionen bei allen Allelen die genau gleichen physikochemischen Differenzen aufweisen. Tabelle III.11 zeigt eine Gegenüberstellung der Ergebnisse aus den Ähnlichkeitsanalysen der Aminosäurepositionen der Matrices 1, 2 und 3.

Tabelle III.11: Kombinationen von Positionen, Assoziationen von Positionen und Positionen, die in Clustern mit langen terminalen Astlängen angeordnet sind. Gezeigt sind die Ergebnisse der ZWHC der Matrices 1 und 2 jeweils nach Spearman'scher Rangkorrelation und nach Bildung der Euklidischen Distanzen und der MDS der Matrices 1 und 3 nach Bildung der Euklidischen Distanzen. In der ersten Spalte links stehen die Positionen, die in den verschiedenen Analysen in Kombinationen (K), Assoziationen (V) oder Clustern (Cl) angeordnet wurden. In Klammern ist angegeben, ob es sich um einzigartige, charakteristische oder allgemeinvariable Positionen handelt. Die Spalte ist unterteilt in Positionen, die bei den Analysen von Matrix 1 (M1) in Kombinationen angeordnet wurden und in Positionen, die den Neuordnungen zugeteilt wurden. Neuordnungen entstanden als Assoziationen und Cluster mit langen terminalen Astlängen bei den Analysen der Matrices 2 und 3 nach Bildung der Euklidischen Distanzen und enthalten Positionen in anderer Zusammensetzung als die bei Analysen von Matrix 1 gebildeten Kombinationen.

Bei der jeweiligen Analyse ist angegeben, ob alle in der ersten Spalte angegebenen Positionen in der Kombination, der Assoziation oder dem Cluster vorhanden (vorh) waren. Waren nicht alle angegebenen Positionen vorhanden, werden die Positionen aufgezählt und in Klammern angegeben, ob es sich um eine Kombination (K), eine Assoziation (Ass) oder ein Cluster (Cl) handelt. Die Angaben für die Analysen von Matrix 1 waren für die Spearman'sche Rangkorrelation und die Bildung der Euklidischen Distanzen identisch und wurden daher in einer Spalte zusammengefasst. Positionen, die Deletionen enthalten, sind kursiv gedruckt.

Positionen der M1-Kombinationen	M1 (ZWHC und MDS)	M2 (ZWHC)		M3 (MDS)
	Spearman / Euklidische Distanzen	Spearman	Euklidische Distanzen	Euklidische Distanzen
7, 23, 24, 54 (A)	vorh (K)	vorh (K)	7, 23 (K)	7, 23 (K)
144, 303, 422 (E)	vorh (K)	303, 422 (K)	-	-
269, 282, 302, 308, 353, 387, 405, 421 (E)	vorh (K)	vorh (K)	[405, 421 (K), 353] (CI) und 302, 308, 387 (Ass) und 269, 282 (CI)	405, 421 (K)
383, 385 (C)	vorh (K)	vorh (K)	(CI)	-
237, 352, 430 (C)	vorh (K)	vorh (K)	-	-
246, 247 (A)	vorh (K)	vorh (K)	-	-
83, 267 (E)	vorh (K)	vorh (K)	vorh (Ass)	-
104, 162, 205, 273, 393 (E)	vorh (K)	vorh (K)	104, 162, 393 (Ass) und 205, 273 (CI)	[162, 393 (K), 104] (Ass)
211, 221, 256, 259, 380 (E)	vorh (K)	221, 256, 259, 380 (K)	[221, 380 (Ass), 259] (CI)	-
57, 265, 307, 370 (E)	vorh (K)	vorh (K)	-	-
305, 356, 358, 417, 418	vorh (K)	305, 356, 418 (K)	-	-
251, 431	vorh (K)	-	-	-
Positionen der Neuordnungen	-	-	-	-
221, 380 (E)	-	-	vorh (Ass)	vorh (Ass)
7, 23, 24, 190, 201, 247 (A)	-	-	[(Ass), 219] (CI)	vorh (Ass)
42, 403 (A)	-	-	vorh (CI)	-
144, 211, 307, 362, 373, 431	-	-	vorh (K)	-
287, 288 (A)	-	-	vorh (CI)	-
321, 322 (A)	-	-	vorh (CI)	-
54, 246, 374 (A)	-	-	vorh (Ass)	vorh (Ass)
57, 173, 390, 418, (E, A)	-	-	vorh (Ass)	57, 173 (Ass)
297, 370 (C, E)	-	-	vorh (CI)	-
256, 230, 372 (E)	-	-	vorh (CI)	-
390, 297 (A, C)	-	-	-	vorh (Ass)
303, 302, 308, 430 (E)	-	-	-	vorh (Ass)
418, 370 (E)	-	-	-	vorh (Ass)
251, 388 (C)	-	-	-	vorh (Ass)
93, 270 (C)	-	-	-	vorh (Ass)

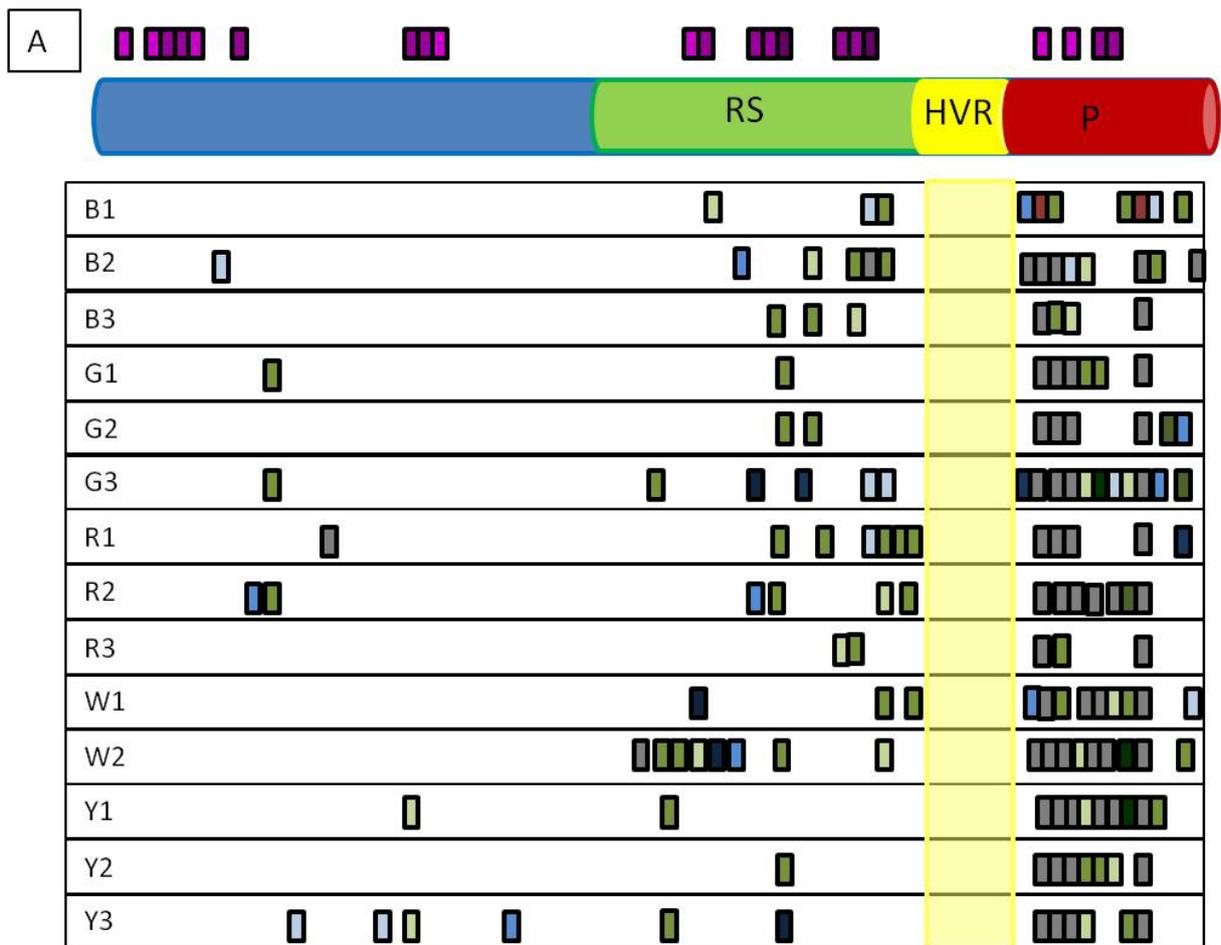
Die physikochemischen Differenzen der Aminosäurepositionen reichen von 5 (Position 372) bis 196 (Position 196). Befinden sich an einer Position mehr als zwei verschiedene Aminosäuren, so treten zwei oder mehr verschiedene physikochemische Differenzen an dieser Position auf. Für die folgende Einteilung der Aminosäurepositionen nach der Höhe ihrer physikochemischen Differenzen wird jeweils der Mittelwert dieser physikochemischen Differenzen verwendet. Es ist zu erkennen, dass einige allgemeinvariable Positionen (7, 23, 54, 219, 246, 374, 390) und einzigartige oder charakteristische Aminosäurepositionen (57 (B2), 104 (Y3), 162 (Y3), 173 (Y1, Y3), 251 (B1, W2), 297 (B2, R3), 302 (G3), 303 (R1), 305 (B1), 308 (G3), 310 (B3, R2, W2, Y1), 368 (W2, Y1, Y3), 370 (B2), 372 (Deletion bei R2, W1, W2, Y1), 387 (G3), 393 (Y3), 418 (B1) und 430 (W1)) geringe physikochemische Differenzen von unter 50 aufweisen. Bei den allgemeinvariablen Positionen, 24, 190, 201, 247, 287, 288, 321, 322, 399 und 403 und bei den einzigartigen oder charakteristischen Positionen 83 (R2), 93 (G1, G3, R2), 205 (Y3), 221 (W2), 230 (G3, W2), 259 (W2), 265 (B2), 267 (R2), 270 (G1, R2, Y2), 281 (B3, G2, R1, W2), 295 (B3, R1, R3), 304 (B2, G2, R1, W1), 312 (B2, R1, W1), 359 (B1, B3, R3, W1), 383 (G1, Y2), 405 (G3), 411 (B1, W1, Y3), 421 (G3), 429 (G2) liegen die physikochemischen Differenzen mit Werten zwischen 51 und 100 im mittleren Bereich.

Die allgemeinvariablen Positionen 293 und 323 und einige einzigartige oder charakteristische Aminosäurepositionen weisen hohe physikochemische Differenzen zwischen 101 und 150 auf. Dies sind die Positionen 282 (G3), 324 (R1, R2), 352 (W1), 353 (G3), 356 (B1), 380 (W2), 385 (G1, Y2), 419 (B1, Y1), 422 (R1), 427 (B2, W2), 428 (G2, G3). Die einzigartigen oder charakteristischen Aminosäurepositionen 237 (W1), 256 (W2), 269 (G3), 273 (Y3) und 376 (G3, R2, W2, Y1) besitzen sehr hohe physikochemische Differenzen zwischen 150 und 196.

Eine Übersicht über die Lokalisation der Aminosäuresubstitutionen und -deletionen der *csd*-Allele und die physikochemischen Differenzen der Aminosäuresubstitutionen zeigt Abbildung III.15. Von den 82 variablen Positionen liegen 66 Positionen (80%) in der RS-Domäne und der prolinreichen Region, vor allem die einzigartigen und die charakteristischen Aminosäuredifferenzen. Die Deletionen, die manche Allele an bestimmten Positionen aufweisen (für genaue Informationen hierzu siehe Tabelle III.10) befinden sich zu 70% in der prolinreichen Region. Die allgemeinvariablen Aminosäurepositionen sind gleichmäßiger über das gesamte *csd*-Gen verteilt als die einzigartigen und die charakteristischen. Es ist erkennbar, dass auch bei den Aminosäurepositionen vor allem jene mit größeren

physikochemischen Differenzen in der RS- Domäne und der prolinreichen Region vertreten sind.

Die Allele B1, B2, G3, R1, R2, W1, W2 und Y3 besitzen jeweils Kombinationen von einzigartigen Positionen. Das Allel G2 hat eine einzigartige Einzelposition und die Allele B3, G1, R3, Y1, Y2 besitzen nur charakteristische Aminosäurepositionen. Durch Einführung der physikochemischen Differenzen zur Bestimmung der Qualitäten der Aminosäureunterschiede in den Matrices 2 und 3 wurde die Einzigartigkeit oder das Charakteristische einiger Aminosäurepositionen für bestimmte Allele deutlich verstärkt oder abgeschwächt. Eine deutliche Abschwächung an mehreren Aminosäurepositionen ist z.B. bei dem Allel Y3 zu beobachten.



█ : ein Balken entspricht einer Aminosäureposition

pcD	A	C	E	
< 50	█	█	█	█ = Deletion
51-100	█	█	█	█ = keine Angabe zur pcD möglich
101-150	█	█	█	
> 150		█	█	

Abbildung III.15: Dargestellt ist die Verteilung von einzigartigen, charakteristischen und allgemeinvariablen Aminosäuresubstitutionen im *csd*-Gen, beruhend auf den Positionswerten der **Matrix 2**. In violett sind die allgemeinvariablen (A) Substitutionen angegeben, in grün die charakteristischen (C) und in blau die einzigartigen (E) Substitutionen. Die Farbintensität der Balken ist ein Maß für die Höhe der physikochemischen Differenz (pcD) an dieser Position. Weist eine Position mehr als eine physikochemische Differenz auf (weil mehr als zwei verschiedene Aminosäuren auftreten) wird der Mittelwert dieser Differenzen dargestellt. Weist ein Allel an einer Aminosäureposition eine Deletion auf, so ist dies durch ein graues Rechteck gekennzeichnet. Können bei einem Allel an einer bestimmten Aminosäureposition keine Angaben zur physikochemischen Differenz gemacht werden, weil alle anderen Allele an dieser Position eine Deletion aufweisen, so ist dies mit einem roten Rechteck gekennzeichnet. Die Aminosäureunterschiede in der hypervariablen Region (gelb unterlegt) wurden in dieser Abbildung nicht berücksichtigt. RS= RS-Domäne; HVR= hypervariable Region; P= prolinreiche Region

III.7 Separate Auswertung der hypervariablen Region (HVR)

Die hypervariablen Regionen der 14 Allele (Aminosäurepositionen 325 bis 351 im Alignment der Allele) können aufgrund vieler Deletionen, variierender Aminosäurezusammensetzungen, unterschiedlicher Anzahl von Repeats und damit unterschiedlicher Länge nicht in einem Alignment angeordnet werden. Sie werden daher unabhängig von den ZWHC und den MDS der Matrices separat ausgewertet. Die Einteilung der hypervariablen Regionen in verschiedene Typen erfolgt gemäß der Aminosäurezusammensetzung der Repeats. Ein Repeat besteht aus Wiederholungen bestimmter Aminosäureabfolgen (Motive). Die Motive können dabei in ihrer Zusammensetzung leicht variieren. Bei den hier untersuchten hypervariablen Regionen handelt es sich um Asparagin-Tyrosin (NY)-Repeats mit variierenden N-Y Anteilen. Bei allen Allelen außer den Allelen B3 und G3 enthalten einer oder zwei der Repeats noch Serin, Cystein, Lysin, Histidin oder Isoleucin.

Die Aminosäuren, die außer Asparagin und Tyrosin in den Repeats der hypervariablen Regionen der verschiedenen *csd*-Allele zu finden sind, machen die HVR des jeweiligen *csd*-Allels typisch. Dies kann durch Betrachtung der physikochemischen Differenzen der Aminosäuren zu den Aminosäuren Asparagin und Tyrosin quantifiziert werden.

Die hypervariablen Regionen der 14 *csd*-Allele habe ich in fünf verschiedene HVR-Typen unterteilt (Abbildung III.16). Dem Typ 1 wurden die hypervariablen Regionen der Allele Y2, G1, G2 und B1 zugeordnet. Sie beginnen mit einer SNYNNY-Sequenz. Darauf folgen bei den verschiedenen Allelen unterschiedlich viele NY-Repeats mit Unterschieden in der N-Y-Zusammensetzung. Die HVR des Allels B1 endet als Besonderheit mit einem Serin. Die physikochemische Differenz von Serin zu Asparagin ist mit 46 relativ gering, die von Serin zu Tyrosin mit 146 hoch. Typ 2 gehören die Hypervariablen Regionen von G3, B3 und Y3 an. Sie zeigen eine geringfügig andere Aminosäurezusammensetzung und Repeatabfolge als die hypervariablen Regionen vom Typ 1 und sind untereinander weniger ähnlich als die hypervariablen Regionen der anderen Typen. Sie beginnen direkt mit N-Y-Repeats unterschiedlicher N-Y-Zusammensetzung und nicht mit S-N-Y. Die HVR des Allels G3 ist sehr kurz (sieben Aminosäuren). Bei Allel Y3 tritt außerdem als Besonderheit ein Lysin (K) im letzten Drittel der HVR auf. Lysin hat zu Asparagin und Tyrosin mittlere physikochemische Differenzen von 85 und 94. Ansonsten unterscheiden sich die hypervariablen Regionen vom Typ 2 nur durch leichte Variationen der Zusammensetzungen der Repeats (unterschiedlich

viele Asparaginreste) und der Anzahl der Repeats. Typ 3 wurden die Allele W2, R3, Y1 und R2 zugeordnet. Deren hypervariablen Regionen beginnen mit 3-4 Asparaginen, gefolgt von einem Y-K-Y-N-Y-Motiv. Darauf folgen relativ unterschiedliche Sequenzen. Die HVR des Allels Y1 endet mit einem Cystein und ist die kürzeste dieses Typs. Die physikochemischen Differenzen von Cystein zu Asparagin und Tyrosin sind mit 139 und 194 hoch. Das Allel R2 besitzt die längste HVR dieses Typs. Sie schließt in ihrem 3'-Bereich ein Lysin in die N-Y-Sequenz ein. Lysin hat zu Asparagin und Tyrosin physikochemische Differenzen von 85 und 94. Die HVR des Allels R3 schließt ebenfalls ein Lysin in ihrem 3'-Bereich ein und endet mit einem Serin. Die HVR des Allels W2 endet mit einer NY-Sequenz und weist in ihrem 3'-Bereich keine Aminosäuren außer Asparagin und Tyrosin auf. Die hypervariablen Regionen der Allele R1 und B2 wurden dem Typ 4 zugeordnet. Sie beginnen mit einer NISNYNNNNN-Sequenz. Die physikochemischen Differenzen von Isoleucin zu Asparagin ist mit 149 hoch, die zu Tyrosin mit 33 sehr niedrig. Die hypervariablen Regionen der beiden Allele unterscheiden sich lediglich in den letzten beiden Aminosäuren (Tyrosin-Asparagin bei R1 und Serin bei B2) voneinander. Die HVR des Allels W1 bildet einen eigenen Typ 5. Sie beginnt mit einem N-N-C-Motiv, besitzt ein 3'-endständiges Histidin und besteht aus nur sieben Aminosäuren. Histidin hat mit 68 und 83 physikochemischen Differenzen von mittlerer Höhe zu Asparagin und Tyrosin. Die HVR des Allels W1 ist mit der HVR des Allels G3 zusammen die kürzeste HVR der Allele.

Y2	SNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNY	Typ 1
G1	SNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNY~	
G2	SNYNNYNNYNNYNNY~	
B1	SNYNNYNNYNNNS~	
G3	NNYNNY~	Typ 2
B3	NNNNYNNNNYNNNNYNNNNYNNNNYNNY~	
Y3	NNNYNNNNYNNNYKNYNS~	
W2	NNNYKYNYNNNNYN~	Typ 3
R3	NNNNYKYNYKNYNS~	
Y1	NNNNYKYNYNNNC~	
R2	NNNNYKYNYNNNNNNYKNYNNY~	
R1	NI SNYNNNNNNYN~	Typ 4
B2	NI SNYNNNNNS~	
W1	NNCNYKH~	Typ 5

Abbildung III.16: Aminosäuresequenzen der hypervariablen Regionen der 14 *csd*-Allele

Die Einteilung der Hypervariablen Regionen in Typen erfolgt gemäß der Aminosäurezusammensetzung der Repeats.

C=Cystein, H=Histidin, I=Isoleucin, K=Lysin, N=Asparagin, S=Serin Y=Tyrosin

Die insgesamt längsten hypervariablen Regionen besitzen die Allele Y2 (27 As), B3 (26 As) und G1 (25 As). Die hypervariablen Regionen des gleichen Typs weisen oft Längenunterschiede von drei und mehr Aminosäuren auf. Die hypervariablen Regionen der Allele R1 und B2 (Typ 4), G2 und B1 (Typ 1) und W2 und R3 (Typ 3), sind Ausnahmen. Sie haben jeweils eine Längendifferenz von nur einer Aminosäure. Die hypervariablen Regionen der Allele R3 und Y1 (Typ 3) haben eine Längendifferenz von zwei Aminosäuren.

III.8 Vergleich der Typeinteilung der hypervariablen Regionen mit den auf den ZWHC und den MDS basierenden Ähnlichkeitsdarstellungen der *csd*-Allele

Um zu beurteilen, ob *csd*-Allele mit ähnlichen hypervariablen Regionen sich auch in den anderen Regionen des *csd*-Gens ähneln, wurde die Einteilung der hypervariablen Regionen in Typen der Ähnlichkeit der Allele, die sich bei den Zwei-Weg hierarchischen Clusteranalysen und den Multidimensionalen Skalierungen zeigte, gegenüber gestellt. Als ähnlich wurden die hypervariablen Regionen eingestuft, die nach Abbildung III.16 einem Typ angehören.

Die hypervariablen Regionen der Allele B1, G1, G2 und Y2 wurden dem Typ 1 zugeordnet. Wobei die hypervariablen Regionen der Allele B1 und G2 deutlich kürzer sind als die der Allele G1 und Y2. Es wurde bei den paarweisen Vergleichen dieser Allele die funktionelle Verschiedenheit nachgewiesen, bis auf die Allele B1 und G1, die im Experiment nicht

kombiniert wurden. Die Allele G1, G2 und Y2 wurden in den Zwei-Weg hierarchischen Clusteranalysen und den Multidimensionalen Skalierungen der Matrices 1, 2 und 3 als sehr ähnlich dargestellt. Das Allel B1 wurde bei den Analysen der Matrix 1 als relativ unähnlich zu den Allelen G1, G2 und Y2 dargestellt. In den Analysen der Matrices 2 und 3 nach Bildung der Euklidischen Distanzen ist B1 den Allelen G1, G2 und Y2 etwas ähnlicher. Die hypervariablen Regionen der Allele B3, G3 und Y3 wurden Typ 2 zugeordnet. In den Zwei-Weg hierarchischen Clusteranalysen und den Multidimensionalen Skalierungen der drei Matrices wurden diese drei Allele als untereinander nicht sehr ähnlich dargestellt. Das Allel B3 wurde experimentell nicht mit dem Allel Y3 kombiniert, so dass die funktionelle Verschiedenheit dieser beiden Allele nicht belegt ist. Dem Typ 3 gehören die hypervariablen Regionen der Allele R2, R3, W2 und Y1 an. Bei den Zwei-Weg hierarchischen Clusteranalysen und den Multidimensionalen Skalierungen wurden diese Allele als relativ ähnlich dargestellt, bis auf das Allel W2. Die hypervariablen Regionen der Allele B2 und R1 sind sich sehr ähnlich und wurden Typ 4 zugeordnet. In den Zwei-Weg hierarchischen Clusteranalysen und den Multidimensionalen Skalierungen wurden sie als nicht sehr ähnlich dargestellt. Experimentell wurden sie nicht kombiniert und so ihre funktionelle Verschiedenheit auch nicht nachgewiesen. Das heißt, dass nicht davon ausgegangen werden kann, dass die hypervariablen Regionen dieser beiden Allele potentiell allelische Spezifität erzeugen können. Die Zusammensetzung der HVR des Allels W1 unterscheidet sich von der aller anderen Allele. In den Zwei-Weg hierarchischen Clusteranalysen und den Multidimensionalen Skalierungen stellte sich heraus, dass das Allel W1 allen anderen Allelen unähnlich ist.

Es lässt sich erkennen, dass Allele, die in den Zwei-Weg hierarchischen Clusteranalysen und den Multidimensionalen Skalierungen als ähnlich dargestellt wurden, häufig, aber nicht immer, auch ähnliche hypervariable Regionen besitzen, so dass kein Zusammenhang zwischen dem HVR-Typ und der Ähnlichkeit der Allele nach Zwei-Weg hierarchischer Clusteranalysen und den Multidimensionalen Skalierungen festgestellt werden kann.

III.9 Ausführlicher Vergleich der Allele G1, G2 und Y2

Die drei Allele G1, G2 und Y2 wurden in den meisten Analysen als sehr ähnlich dargestellt. Die Allele G1 und Y2 sind zudem mit hoher Wahrscheinlichkeit phylogenetisch eng miteinander verwandt (Abbildung III.5). Ihre Aminosäuresequenzen unterscheiden sich

außerhalb der HVR nur an sechs Aminosäurepositionen (7, 23, 24, 42, 54 und 93) und haben somit von allen Allelen die wenigsten Aminosäuredifferenzen. Die Ursache für die funktionelle Verschiedenheit dieser beiden Allele kann daher auf die Aminosäurepositionen 7, 23, 24, 42, 54 und 93 und die HVR eingegrenzt werden. Die Positionen 7, 23, 24 und 54 bilden laut ZWHC und MDS-Analysen nach Spearman'scher Rangkorrelation eine Kombination allgemeinvariabler Aminosäurepositionen und Position 42 ist eine allgemeinvariable Einzelposition. Die physikochemischen Differenzen dieser Positionen nehmen Werte zwischen 29 (Position 54) und 64 (Position 42) an. Position 93 ist charakteristisch für die Allele G1, G3 und R2 und hat eine physikochemische Differenz von 59. Die hypervariablen Regionen der beiden Allele sind sehr ähnlich. Sie unterscheiden sich lediglich im 3' Bereich des N-Y-Repeats. Die HVR des Allels Y2 ist dabei um drei Aminosäuren länger als die des Allels G1.

Die Allele G2 und Y2 unterscheiden sich an 16 Aminosäurepositionen in der Region zwischen den Aminosäurepositionen 190 und 429. Von diesen bilden zwei Positionen (246 und 247) laut ZWHC und MDS-Analysen nach Spearman'scher Rangkorrelation eine Kombination allgemeinvariabler Aminosäurepositionen und zwei Positionen (383 und 385) sind charakteristisch für die Allele G1 und Y2. Eine Position (429) ist einzigartig für das Allel G2. Die übrigen Positionen sind für drei bzw. vier Allele charakteristisch oder allgemeinvariabel. Die physikochemischen Differenzen liegen fast alle zwischen 46 (Position 219) und 95 (Position 281). Position 293 hat mit 125 die größte physikochemische Differenz. Die hypervariablen Regionen der beiden Allele unterscheiden sich durch eine große Längendifferenz von 13 Aminosäuren.

Abbildung III.17 zeigt die Übersicht über die Verteilung der Aminosäuredifferenzen der Allele G1 und G2 zu dem Allel Y2. Während die Aminosäuredifferenzen von G1 zu Y2 am 5'-Ende des *csd*-Gens liegen, befinden sich die Unterschiede der Allele G2 und Y2 im Bereich der Mitte des Gens bis zum 3'-Ende. Es gibt keine Aminosäurepositionen an denen sie die Allele G1 und G2 beide gemeinsam von dem Allel Y2 unterscheiden.

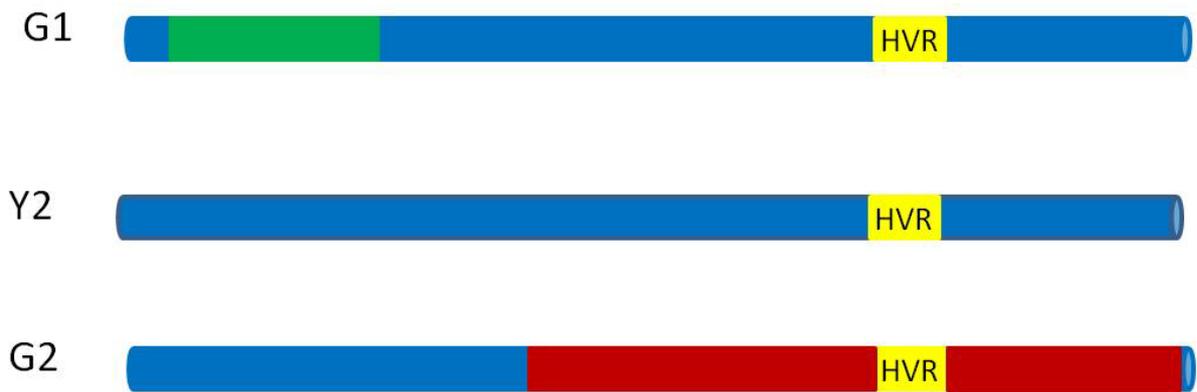


Abbildung III.17: Verteilung der Aminosäuredifferenzen der *csd* Allele G1 und G2 zu Allel Y2.

Die Regionen an denen das allel G1 Aminosäureunterschiede zu dem Allel Y2 aufweist sind grün markiert, die Regionen, in denen das Allel G2 zu dem Allel Y2 Aminosäureunterschiede aufweist, sind rot. Die hypervariable Region ist unabhängig von allelelischer Variabilität gelb dargestellt.

IV. Diskussion

IV.1 Die experimentell kombinierten *csd*-Allele sind funktionell verschieden

In der Natur segregieren bis zu 19 funktionell verschiedene *csd*-Allele in einer Population (Adams et al. 1977). Über die Spezifität von *csd*-Allelen ist noch nicht viel bekannt. Sie ist in den Aminosäuresequenzen der Allele kodiert. Um Aminosäureunterschiede zwischen funktionell verschiedenen *csd*-Allelen untersuchen zu können, wurden in einem Kreuzungsexperiment 15 *csd*-Allele darauf getestet, ob sie funktionell verschieden sind. In fünf Parental (P)- Kreuzungen wurden fünf P-Königinnen mit den Spermien von einem von fünf P-Drohnen inseminiert. In der ersten Filial (F1)-Generation wurden 23 F1-Königinnen mit den Spermien von F1-Drohnen besamt. Die weiblichen Nachkommen der F1-Generation (F2-Arbeiterinnen) wurden genotypisiert, um zu ermitteln, welche paarweise kombinierten *csd*-Allele in Weibchen existieren und somit funktionell verschieden sind. Die Zygoten, in denen die verschiedenen *csd*-Allele paarweise vereint wurden, entwickelten sich zu Weibchen daher sind alle *csd*-Allele, die in dem Kreuzungsexperiment paarweise kombiniert wurden, funktionell verschieden.

Bisher ist bekannt, dass diploide homozygote Individuen, die sich zu männlichen Bienen entwickeln würden, nach dem Schlupf von Arbeiterinnen gefressen werden (Mackensen, 1955). Fünf der 23 F1-Kreuzungen sind Inzuchtkreuzungen, in denen F1-Königinnen mit Spermien ihrer Geschwisterdrohnen besamt wurden. Die Nachkommen dieser Inzuchtkreuzungen sind zu 50% homozygot. In den weiblichen Nachkommen (F2-Arbeiterinnen) dieser Inzuchtkreuzungen wurden die homozygoten Allele erwartungsgemäß nicht gefunden. In einigen F2-Eiern der Inzuchtkreuzungen konnte jeweils nur ein *csd*-Allel identifiziert werden. Dieses Allel entsprach in allen Fällen dem zu erwartenden homozygoten Allel. Mit großer Wahrscheinlichkeit handelt es sich daher bei diesen Funden um homozygote diploide Individuen. Somit wäre gezeigt, dass homozygote *csd*-Allele in den befruchteten Eiern existieren, in den adulten Bienen aber nicht mehr.

Es besteht die Möglichkeit, dass parentales Imprinting einen Einfluss auf Spezifität von *csd*-Allelen haben könnte. Bei der halodiploiden Wespe *Nasonia vitripennis* wird z.B. vermutet dass parentales Imprinting bei der Geschlechtsdetermination eine bedeutende Rolle spielt. Parentales Imprinting führt bei dieser Spezies vermutlich dazu, dass biparentale Zygoten sich zu Weibchen entwickeln und uniparentale zu Männchen (Trent et al. 2006). Um den

möglichen Einfluss von parentalem Imprinting auf die Spezifität von *csd*-Allelen der Honigbiene zu untersuchen, wurden einige Allele in dem Kreuzungsexperiment reziprok kombiniert. Alle reziprok kombinierten *csd*-Allele wurden in den F₂-Arbeiterinnen gefunden. Die weibliche Entwicklung von Zygoten wird also nicht davon beeinflusst, welches der beiden *csd*-Allele vom Drohnen stammt und welches von der Königin. Es konnte somit kein Einfluss von paternalem Imprinting auf die Spezifität von *csd*-Allelen nachgewiesen werden.

IV.2 Außer bei den Allelen G1 und Y2 konnte keine phylogenetische Verwandtschaft der *csd*-Allele nachgewiesen werden

Mit Ausnahme der Allele G1 und Y2 konnte keine gemeinsame Abstammung der *csd*-Allele festgestellt werden. Das könnte bedeuten, dass die *csd*-Allele, mit Ausnahme der Allele G1 und Y2, vor sehr langer Zeit divergiert sind. Während großer Zeiträume konnten Aminosäuresubstitutionen bei den verschiedenen Allelen mehrfach entstehen und auch wieder verschwinden. Die Anzahl, die Reihenfolge und die Positionen der Aminosäuresubstitutionen änderte sich bei den Allelen in diesem Zeitraum so oft, dass nicht mehr erkennbar ist, welche der Allele ursprünglich ähnlich waren. Daher ist keine statistisch abgesicherte Aussage über die phylogenetische Verwandtschaft möglich.

Die Auswertung der Aminosäureunterschiede kann daher unabhängig von phylogenetischer Verwandtschaft erfolgen. Die Allele G1 und Y2 bilden die Ausnahme. Sie sind mit großer Wahrscheinlichkeit phylogenetisch eng verwandt. Die Aminosäureunterschiede dieser beiden Allele wurden, zusätzlich zu den Analysen mittels ZWHC und MDS, in Abschnitt IV.13 separat diskutiert.

IV.3 Welche Regionen des *csd*-Gens könnten für die Spezifität der Allele besonders wichtig sein?

Bei den Analysen der 14 *csd*-Allele in dieser Arbeit wurden die Allele in einem Alignment angeordnet. Das Alignment bestand aus 431 Aminosäurepositionen von denen 82 Positionen Aminosäuredifferenzen und / oder Deletionen aufwiesen. Bei den Analysen mittels Zwei-Weg hierarchischer Clusteranalyse (ZWHC) und Multidimensionaler Skalierung (MDS) zeigte

sich, dass sich die variablen Aminosäurepositionen unterteilen lassen in einzigartige, charakteristische und allgemeinvariable Positionen. Bei den einzigartigen Positionen ist die Aminosäure jeweils nur bei einem Allel verändert, bei den charakteristischen Positionen ist die Aminosäure jeweils an zwei bis vier Allelen verändert und bei den allgemeinvariablen Positionen ist die Aminosäure bei mehr als vier Allelen verändert.

Es ist bekannt, dass das *csd*-Gen unter balancierender Selektion steht (Hasselmann et al, 2008 (1); Charlesworth, 2004; Hasselmann and Beye, 2004). Hasselmann und Mitarbeiter (2008(1)) konnten das genaue Ziel der balancierenden Selektion innerhalb des *csd*-Gens auf die Exone 6 und 7 eingrenzen. Hier wurden die meisten synonymen Substitutionen innerhalb des *csd*-Gens gefunden. Es wurde postuliert, dass diese Region die potentiell spezifizierende Domäne (PSD) des *csd*-Gens kodiert. Die RS-Domäne und die HVR liegen zum großen Teil in diesen beiden Exonen (Übersicht über die Exone und Domänen bzw. Regionen in Abbildung I.2). Wie in Abbildung III.15. erkennbar, sind die einzigartigen und die charakteristischen Aminosäuresubstitutionen und -deletionen außerhalb der HVR hauptsächlich in der RS-Domäne und der prolinreichen Region der Allele zu finden. Die prolinreiche Region beinhaltet außerdem 70% der beobachteten Deletionen. Die allgemeinvariablen Positionen sind verhältnismäßig gleichförmig über das *csd*-Gen verteilt. Insgesamt weisen einzigartige, charakteristische und allgemeinvariable Aminosäurepositionen sowohl hohe als auch niedrige physikochemische Differenzen auf, wobei bei den einzigartigen und den charakteristischen Positionen hohe physikochemische Differenzen (>100) deutlich öfter auftraten als bei den allgemeinvariablen Positionen (dreizehnmal im Vergleich zu zweimal). Bei den einzigartigen Positionen wurden außerdem auch die absolut größten physikochemischen Differenzen gefunden (Positionen 237, 256 und 269 mit physikochemischen Differenzen von 196, 175 und 181). Tendenziell sind die physikochemischen Differenzen der einzigartigen, charakteristischen und allgemeinvariablen Positionen in der RS-Domäne und der prolinreichen Region höher als in den Exonen 2-5, die C-terminal der RS-Domäne liegen. In bisherigen Analysen der Aminosäurepolymorphismen bei *csd*-Allelen (Beye et al., 2003; Hasselmann and Beye, 2004; Hasselmann and Beye, 2006; Hasselmann et al., 2008 (1)) wurden Aminosäurepositionen, die Deletionen aufweisen, nicht in die Analysen einbezogen und damit auch nicht die durch Deletionen verursachte Variabilität. Bei den von mir durchgeführten Zwei-Weg hierarchischen Clusteranalysen und Multidimensionalen Skalierungen der Matrix 1 wurden Aminosäurepositionen, die

Deletionen aufweisen, in die Analyse einbezogen soweit Positionswerte in den Matrices vorhanden waren und außerdem in einer separaten Analyse bewertet (Abschnitt IV.7). Da in der prolinreichen Region 70% der Deletionen zu finden sind, wurde vor allem die Variabilität dieser Region durch die Berücksichtigung der Deletionen erhöht. Der potentielle Beitrag von Deletionen zur Spezifität von *csd*-Allelen wird in Abschnitt IV.7 diskutiert. Die Aminosäuresubstitutionen und -deletionen sind zahlenmäßig in der RS- Domäne und der prolinreichen Region konzentriert und haben dabei tendenziell höhere physikochemischen Differenzen als in der Exon 2-5 Region. Dies könnte darauf hinweisen, dass die RS- Domäne und die prolinreiche Region besondere Bedeutung für die Spezifität von *csd*-Allelen haben. Die Bedeutung der RS-Domäne und der prolinreichen Region für die in der Einleitung vorgestellten Proteinbindungsmodelle (Abschnitt I.3) wird in Abschnitt IV.12 diskutiert.

Im C-terminalen Bereich (Exon 2-5 Region) des *csd*-Gens liegen nur 20 % der variablen Positionen. In der Exon 2-5 Region von *csd*-Allelen wurden sporadische Rekombinationsereignisse gefunden (Hasselmann et al., 2008 (1)). Rekombination hat homogenisierende Wirkung auf die Regionen, auf die sie wirkt. Spezifität, die durch das Zusammenspiel bestimmter Aminosäuresubstitutionen kodiert ist, würde so verloren gehen. Durch Rekombination werden die Exone 2-5 dem Einfluss der balancierenden Selektion entzogen, die auf die putative PSD von *csd* (Exone 6 und 7) wirkt. Das könnte bedeuten, dass die Exone 2-5 von nicht so großer Bedeutung für die Spezifität von *csd*-Allelen sind, wie die Exone 6 und 7 (Hasselmann et al., 2008 (1)). Es wäre möglich, dass Aminosäuresubstitutionen im *csd*-Gen außerhalb der PSD nur geduldet sind oder der Aufrechterhaltung der Proteinkonformation dienen. Bei der Kinase Domäne der Serin-Threonin Rezeptor Kinase (SRK) der Pflanze *Arabidopsis lyrata* wurden zum Beispiel Polymorphismen entdeckt, die vermutlich keine Bedeutung für die Spezifität des Proteins haben (Charlesworth, 2003). *Arabidopsis lyrata* besitzt ein sporophytisches Selbstinkompatibilitätssystem. Im S-Lokus dieser Pflanze befindet sich unter anderem das Gen für die SRK, die aus einer extrazellulären S-Domäne und aus einer intrazellulären Kinasedomäne besteht. Die S-Domänen der SRK-Haplotypen unterscheiden sich in bis zu 30% ihrer Aminosäuren (Charlesworth, 2003). Die S- Domäne ist vermutlich für die Spezifität der SRK verantwortlich. In der Kinase- Domäne wurden weniger Polymorphismen gefunden. Diese Domäne hat vermutlich nur eine Funktion in der Signalkaskade. Damit diese Funktion aufrecht erhalten werden kann, werden nur Aminosäuresubstitutionen akzeptiert, die die

Funktion nicht beeinträchtigen. Die in der Kinase-Domäne beobachteten Polymorphismen sind daher vermutlich nicht relevant für die Spezifität des Proteins und es handelt sich um neutrale oder schwach selektierte Varianten (Charlesworth, 2003). Die *celA* Gene von Reis (*Oryza sativa*) und Baumwolle (*Gossypium hirsutum*) sind Homologe des bakteriellen *celA* Gene (*Acetobacter xylinum*), die die katalytische Untereinheit der Zellulosesynthase kodieren. Der C- und N-Terminus und hypervariable Regionen dieser Gene weisen hohe Variabilität auf. Die hochkonservierten Bereiche zwischen den hypervariablen Regionen sind vermutlich für Katalyse und / oder Substratbindung (UDP-Glucose) verantwortlich. In diesen Regionen wurden ebenfalls einige Polymorphismen beobachtet, die aber die Aminosäuren nicht betreffen, die für die putative Funktion (Bindung von UDP-Glucose) bedeutsam sind (Pear et al., 1996).

Die Aminosäuresubstitutionen der stärker konservierten Bereiche der Exon 2-5 Region von *csd* könnten wichtig sein für die Spezifität der Allele. Die Substitutionen könnten aber auch, wie es bei der Kinase-Domäne der SRK von *A. lyrata* und den konservierten Bereichen der *celA* Gene von Baumwolle und Reis vermutet wird, neutrale oder schwach selektierte Substitutionen sein, die die Funktion der Csd-Proteine und / oder deren Spezifität nicht beeinflussen.

IV.4 Die Ergebnisse der Zwei-Weg hierarchischen Clusteranalyse und der Multidimensionalen Skalierung können in Abhängigkeit von unterschiedlichen Hypothesen interpretiert werden

Die Beurteilung der Bedeutung der variablen Positionen für die Spezifität von *csd*-Allelen kann in Abhängigkeit von verschiedenen Hypothesen erfolgen. Es wäre möglich, dass alle Aminosäuresubstitutionen potentiell die gleiche Bedeutung für Spezifität haben, und Spezifität kodiert wird durch das bloße Vorhandensein von Variabilität (Aminosäuresubstitutionen und -deletionen) (Hypothese 1). In Matrix 1 wurden Aminosäuresubstitutionen und -deletionen entsprechend bewertet. Hypothese 1 lässt sich in zwei Hypothesen (1a und 1b) unterteilen. Bei Hypothese 1a ist es für die potentielle Bedeutung einer Aminosäureposition für die Spezifität der Allele wichtig, wie hoch der Positionswert des Allels an der betrachteten Position im Vergleich zu den Positionswerten

der anderen Allele an der Position ist. Die Positionswerte der Allele werden nach ihrer Höhe absteigend angeordnet und den Allelen entsprechend Ränge für die betrachtete Position zugewiesen. Das Allel mit dem höchsten Positionswert bekommt dabei den höchsten Rang für die Position, das Allel mit dem niedrigsten Positionswert, den niedrigsten Rang. Die absoluten Positionswerte der Allele spielen daher bei der Beurteilung der potentiellen Bedeutung einer Aminosäureposition für die Spezifität von Allelen nach Hypothese 1a keine Rolle, sondern nur die Ränge der Allele an der Position. Je höher der Rang eines Allels bei einer Aminosäureposition ist, desto größer ist der relative Wert des Allels für die betrachtete Position. Ist der relative Wert eines Allels für eine Position groß, so ist diese Position für das Allel charakteristisch oder sogar einzigartig. Positionen mit unterschiedlichen relativen Werten der Allele sind für unterschiedliche Allele charakteristisch oder einzigartig oder weisen unterschiedliche allgemeine Variabilität auf. Mittels der Spearman'schen Rangkorrelation wurden die Ränge, die den Allelen bei den verschiedenen Positionen zugewiesen wurden, bei allen paarweisen Vergleichen der Positionen korreliert und so die Qualität des linearen Zusammenhangs der Ränge ermittelt. Mittels der Analysen von Matrix 1 nach Spearman'scher Rangkorrelation der Ränge, die die Allele bei den verschiedenen Positionen haben, lassen sich daher Aussagen darüber machen, wie ähnlich die relativen Werte der Allele für die verschiedenen Positionen unter Hypothese 1a sind.

Bei Hypothese 1b sind die absoluten Positionswerte, die die Allele an einer Position aufweisen, wichtig dafür, ob die Position potentiell Bedeutung für die Spezifität von *csd*-Allelen hat. Da in Hypothese 1b der Wert eines Allels für eine Position von den absoluten Positionswerten der Allele an der betrachteten Position abhängt, wird dieser Wert als absoluter Wert eines Allels für eine Position bezeichnet. Je größer der Positionswert eines Allels bei einer Aminosäureposition ist, desto größer ist der absolute Wert des Allels für die betrachtete Position. Ist der absolute Wert eines Allels für eine Position groß, so ist diese Position für das Allel charakteristisch oder sogar einzigartig. Mittels der Bildung der Euklidischen Distanzen der Positionen aus Matrix 1 lässt sich die Ähnlichkeit der Positionen nach Hypothese 1b analysieren. Die Ähnlichkeit der absoluten Werte, die die Allele an den verschiedenen Positionen haben, wurde durch paarweise Bildung der Euklidischen Distanzen der Positionswerte der Allele an jeweils zwei zu vergleichenden Positionen ermittelt. Je unterschiedlicher die absoluten Werte der Allele an den beiden verglichenen Positionen sind, desto größer ist die Euklidische Distanz der Positionen. Positionen mit großer

Euklidischer Distanz sind für unterschiedliche Allele charakteristisch oder einzigartig oder weisen unterschiedliche allgemeine Variabilität auf.

Es wäre aber auch möglich, dass nicht nur das bloße Vorhandensein von Variabilität für die Kodierung von Spezifität von Bedeutung ist, sondern auch das Ausmaß der physikochemischen Änderung, die durch den Austausch der Aminosäuren entstanden ist (Hypothese 2). Hypothese 2 lässt sich, wie Hypothese 1, unterteilen in die Hypothesen 2a und 2b. Nach Hypothese 2a könnte der relative Wert der Allele für die Positionen ausschlaggebend dafür sein, ob die Positionen für die Spezifität von *csd*-Allelen von Bedeutung sind, nach Hypothese 2b der absolute Wert der Allele für die Positionen. Die relativen Werte der Allele für die Positionen sind unter Berücksichtigung der physikochemischen Differenzen nicht mehr alleine davon abhängig, für welche Allele die Position charakteristisch oder einzigartig ist, wie bei Hypothese 1, sondern auch von den physikochemischen Differenzen, die durch die Aminosäuresubstitutionen erzeugt wurden. Je größer die physikochemische Differenz eines Allels zu den anderen Allelen an der betrachteten Position, desto höher ist der Rang des Allels für diese Position. Die physikochemische Differenz wäre unter Hypothese 2a nur innerhalb der betrachteten Position für die Bildung der Ränge der Allele von Bedeutung. Die Ränge, die die Allele an den Positionen haben, können sich von den Rängen unterscheiden, die nur auf dem bloßen Vorhandensein von Variabilität (Hypothese 1a) vergeben wurden. Mittels der Spearman'schen Rangkorrelation der Matrices 2 und 3 kann die Ähnlichkeit von Aminosäurepositionen nach Hypothese 2a analysiert werden.

Der absolute Wert der Allele für eine Position nach Hypothese 2b ist nicht ausschließlich davon abhängig, von wie vielen Allelen sich ein Allel an der betrachteten Position unterscheidet, wie bei Hypothese 1b, sondern auch davon, wie groß die physikochemischen Differenzen dieser Aminosäureunterschiede sind. Der absolute Wert eines Allels für eine Position, der auf bloßer Variabilität beruht, kann deutlich von dem absoluten Wert abweichen, der auf den physikochemischen Differenzen beruht. Hat eine einzigartige Aminosäureposition, bei der die substituierte Aminosäure aber nur eine geringe physikochemische Differenz aufweist, so ist die Euklidische Distanz dieser Position zu einer Position, die zwar für ein anderes Allel einzigartig ist, aber ebenfalls nur eine geringe physikochemische Differenz aufweist, gering. Mittels Bildung der Euklidischen Distanzen der Matrices 2 und 3 kann die Ähnlichkeit der Aminosäurepositionen unter der Hypothese 2b

analysiert werden. Es kann festgestellt werden, welche Allele Positionen besitzen, die sich durch besonders große physikochemische Differenzen von den anderen Positionen abheben und welche Positionen sich durch niedrige physikochemische Differenzen kaum voneinander unterscheiden, obwohl sie für unterschiedliche Allele einzigartig oder charakteristisch sind oder unterschiedliche allgemeine Variabilität aufweisen. Die Ergebnisse der Positionsanalysen mittels ZWHC und MDS werden im Folgenden in Abhängigkeit der verschiedenen Hypothesen interpretiert.

IV.5 Welche Aminosäurepositionen könnten potentiell wichtig für die Spezifität von *csd*-Allelen sein?

Bei den Analysen von Matrix 1 nach Spearman'scher Rangkorrelation und nach Bildung der Euklidischen Distanzen wurden insgesamt zwölf Kombinationen von Positionen gebildet, so dass sich von den 82 variablen Positionen knapp 50% (38 Positionen) in diesen Kombinationen befanden. Positionen, die sich in Kombinationen befinden, wurden bei der ZWHC und der MDS als gleich dargestellt. Von den zwölf Kombinationen bestehen zwei aus allgemeinvariablen Aminosäurepositionen, acht aus einzigartigen Positionen und zwei aus charakteristischen Positionen. Die 44 der 82 Positionen, die nicht in Kombinationen eingeordnet wurden, wurden als allgemeinvariable, einzigartige oder charakteristische Einzelpositionen dargestellt (Tabellen III. 4 und III.5). Die Analysen nach Spearman'scher Rangkorrelation und nach Bildung der Euklidische Distanzen lieferten gleiche Ergebnisse bezüglich der Bildung der Kombinationen von allgemeinvariablen, einzigartigen und charakteristischen Positionen. In Matrix 1 sind die Euklidischen Distanzen proportional zur Anzahl der Unterschiede (Aminosäuresubstitutionen oder -deletionen), da alle Unterschiede den gleichen Wert bekamen. Daher ist der relative Wert der Allele für die Positionen gleich dem absoluten Wert. Die Analyseergebnisse werden daher nicht getrennt nach Hypothese 1a und Hypothese 1b, sondern zusammenfassend unter Hypothese 1 interpretiert.

Die Allele B1, B2 G3, R1, R2, W1, W2 und Y3 besitzen Kombinationen von einzigartigen Aminosäurepositionen (Tabelle III.5). Bei dem Allel G2 wurde eine einzigartige Einzelposition

gefunden und die Allele B3, G1, R3, Y1 und Y2 besitzen keine einzigartigen Positionen, sondern nur charakteristische (Tabellen III.4 und III.5).

Die Positionen, die zusammen in eine Kombination eingeteilt wurden, könnten gemeinsam die Spezifität von *csd*-Allelen beeinflussen. Die einzigartigen Positionen verleihen den jeweiligen Allelen Einmaligkeit und sind daher potentiell besonders wichtig für die Spezifität der jeweiligen *csd*-Allele. Dass bei den Allelen B3, G1, R3, Y1 und Y2 keine einzigartigen Positionen gefunden wurden, lässt allerdings vermuten, dass einzigartige Positionen nicht unbedingt notwendig sind, um Spezifität hervorzurufen. Die allgemeinvariablen Positionen, die sich in Kombinationen befinden, weisen gemeinsam die gleiche allgemeine Variabilität (Aminosäuresubstitutionen bei den gleichen Allelen) auf.

Die Zusammensetzung der Kombinationen allgemeinvariabler, einzigartiger und charakteristischer Positionen wurde durch Berücksichtigung der physikochemischen Differenzen (Matrices 2 und 3) nach Spearman'scher Rangkorrelation nicht verändert. Es kann daher davon ausgegangen werden, dass sich der relative Wert der Allele für die Positionen durch Berücksichtigung der physikochemischen Differenzen (Hypothese 2a) nicht geändert hat.

Wird von Hypothese 2b ausgegangen, so werden die Analyseergebnisse der Matrices 2 und 3 nach Bildung der Euklidischen Distanzen interpretiert. Die Zusammensetzung der Kombinationen allgemeinvariabler, einzigartiger und charakteristischer Positionen hat sich in diesen Analysen grundlegend geändert (Tabelle III.11). Sowohl bei der ZWHC von Matrix 2 als auch bei der MDS von Matrix 3 wurden nur Positionen, die exakt die gleichen Positionswerte bei den verschiedenen Allelen aufweisen (Positionen, die jeweils für das gleiche Allel einzigartig oder für die gleichen Allele charakteristisch sind oder gemeinsam die gleiche allgemeine Variabilität aufweisen und dabei genau gleiche physikochemische Differenzen haben) einer Kombination zugeordnet. Positionen, die zwar für die gleichen Allele einzigartig oder charakteristisch sind, aber unterschiedliche physikochemische Differenzen aufweisen, wurden je nach Höhe der physikochemischen Differenzen Assoziationen von Positionen (ZWHC und MDS) oder Clustern mit langen terminalen Astlängen (ZWHC) zugeordnet. Bei großen physikochemischen Differenzen wurden sie als einander unähnlich dargestellt. Assoziationen von Positionen bestehen entweder aus Positionen, die für die gleichen Allele einzigartig oder charakteristisch sind bzw. die gleiche allgemeine Variabilität besitzen aber geringe Unterschiede in ihren physikochemischen

Differenzen aufweisen, oder aus Positionen, die für unterschiedliche Allele einzigartig oder charakteristisch sind bzw. unterschiedliche allgemeine Variabilität besitzen, aber nur geringe physikochemische Differenzen aufweisen. Cluster mit langen terminalen Astlängen beinhalten Positionen, die für die gleichen Allele einzigartig oder charakteristisch sind bzw. die gleiche allgemeine Variabilität besitzen, aber deutliche Unterschiede in ihren physikochemischen Differenzen aufweisen. Bei der Bildung der Euklidischen Distanzen ist der Wert der physikochemischen Differenzen offensichtlich die entscheidende Kenngröße, die die Ähnlichkeit der Aminosäurepositionen bestimmt. Die Merkmale der Positionen (einzigartig, charakteristisch oder allgemeinvariabel) werden erst an zweiter Stelle für die Bestimmung der Ähnlichkeit berücksichtigt (Cluster mit langen terminalen Astlängen). Doch wie ist der Wert der physikochemischen Differenz einer Aminosäuresubstitution in Bezug auf deren Einfluss auf die Spezifität von *csd*-Allelen einzustufen?

Der erste Schritt jeder Proteinwirkung ist die Bindung eines anderen Moleküls (Stryer, 1995; Alberts et al., 1998). Diese Bindung zeigt immer große Spezifität (Alberts et al. 1998). Die Spezifität von *csd*-Allelen wird könnte vermittelt werden über die Bindungsspezifität der Csd-Proteine. Diese Annahme würde die, in der Einleitung vorgestellten (Abschnitt I.3), hypothetischen Proteinbindungsmodelle 1 und 2 unterstützen, bei denen davon ausgegangen wird, dass die Spezifität der *csd*-Allele durch spezifische Bindungseigenschaften der Csd-Proteine zustande kommt. Der Bezug der Analyseergebnisse zu den Proteinbindungsmodellen wird in Abschnitt IV.12 diskutiert.

Die Fähigkeit eines Proteins, selektiv und mit hoher Affinität ein anderes Molekül zu binden basiert auf der Bildung von schwachen, nicht kovalenten Bindungen (Wasserstoff-Brücken, Ionische Bindungen, van-der-Waals-Interaktionen) und hydrophoben Wechselwirkungen, die durch die physikochemischen Eigenschaften der Aminosäuren ermöglicht werden (Alberts et al., 1998). Auch die Konformation von Proteinen wird durch Wasserstoff-Brücken, Ionischen Bindungen, van-der-Waals-Interaktionen, Disulfid-Brücken und hydrophoben Wechselwirkungen stabilisiert (Alberts et al., 1998, Stryer, 1995). Konformationsänderungen des Proteins können dessen Bindungsspezifität ändern (Alberts et al., 1998). Miyata et al. (1979) haben festgestellt, dass Aminosäuresubstitutionen in Proteinen, die große physikochemische Differenzen verursachen, die Konformation von Proteinen viel stärker beeinflussen können, als Substitutionen, die niedrige physikochemische Differenzen verursachen. Stone and Sidow (2005) berichteten, dass die Beeinträchtigung der

Proteinfunktion mit den physikochemischen Differenzen der substituierten Aminosäuren korreliert ist. Wird eine Aminosäure durch eine andere Aminosäure ersetzt, die ähnliche physikochemische Eigenschaften, wie die Originalaminosäure aufweist, so bleibt die Funktionsfähigkeit des Proteins eher erhalten als bei Substitution durch eine Aminosäure, die andere physikochemische Eigenschaften als die Originalaminosäure aufweist (Schmitt et al., 2009). Campo und Mitarbeiter (2008) fanden heraus, dass es Positionen in Proteinen gibt, an denen geringe Abweichungen von der physikochemischen Eigenschaft der Originalaminosäure toleriert werden. Die Erkenntnisse von Myiata, Schmitt und Campo beziehen sich auf Proteine oder Domänen von Proteinen, in denen die physikochemischen Eigenschaften konserviert sein müssen, um die Funktionsfähigkeit und die Bindungseigenschaften des Proteins aufrecht zu erhalten (negative Selektion). Die PSD von *csd* ist aber auch das Ziel balancierender Selektion (Hasselmann et al, 2008 (1)). Sie wird auch daraufhin selektiert, dass bestimmte Aminosäureunterschiede fixiert werden. Es ist daher vermutlich von Bedeutung, dass die physikochemischen Differenzen dieser Aminosäureunterschiede so hoch sind, dass sie Auswirkungen auf die Bindungseigenschaften des Proteins haben. Die Bindungsspezifität von Csd-Proteinen könnte direkt durch Aminosäuresubstitutionen an Positionen, die in die Bindung eines anderen Proteinmoleküls involviert sind, erzeugt werden, ohne dass die Proteinkonformation verändert wird. Aber auch eine Änderung der Proteinkonformation, die durch Aminosäuresubstitutionen erreicht wird, könnte sich auf die Bindungsspezifität der Csd-Proteine auswirken. Welche Positionen hinsichtlich der Werte ihrer physikochemischen Differenzen potentiell höhere bzw. geringere Bedeutung für die Bindungsspezifität von Csd-Proteinen haben könnten wird nun im Folgenden erörtert.

Die Zwei-Weg hierarchische Clusteranalyse von Matrix 2 und die Multidimensionale Skalierung von Matrix 3 zeigen, dass sieben allgemeinvariable Positionen, sechs charakteristische und zwölf einzigartige Positionen niedrige physikochemische Differenzen von unter 50 besitzen (Tabelle III.6 und Abbildung III.15), so dass die Variabilität dieser Positionen nach Berücksichtigung der physikochemischen Differenzen kaum noch erkennbar ist. Aufgrund dieser niedrigen physikochemischen Differenzen haben diese Positionen nach Hypothese 2b keinen oder nur wenig Einfluss auf die Bindungsspezifität der Csd-Proteine. Es könnte sich um Mutationen handeln, die die Bindungsspezifität nicht signifikant verändern. Bei acht allgemeinvariablen Positionen, zehn charakteristischen und neun einzigartigen

Positionen haben die physikochemischen Differenzen Werte von 51 bis 100. Die Substitutionen an diesen Positionen haben potentiell mehr Einfluss auf die Bindungsspezifität als die Substitutionen an den Positionen mit physikochemischen Differenzen unter 50. Die beiden allgemeinvariablen Positionen 293 und 323 haben hohe physikochemische Differenzen, die zwischen 101 und 150 liegen. Es wäre also denkbar, dass diese beiden Positionen die Spezifität von *csd*-Allelen beeinflussen. Die einzigartigen oder charakteristischen Positionen 282 (G3), 324 (R1, R2), 352 (W1), 353 (G3), 356 (B1), 380 (W2), 385 (G1, Y2), 419 (B1, Y1), 422 (R1), 427 (B2, W2), 428 (G2, G3) besitzen ebenfalls physikochemische Differenzen zwischen 101 und 150, so dass diese Positionen ebenfalls zur Spezifität der *csd*-Allele beitragen könnten. Die höchsten physikochemischen Differenzen (zwischen 151 und 196) weisen die einzigartigen oder charakteristischen Aminosäurepositionen 237 (W1), 256 (W2), 269 (G3), 273 (Y3) und 376 (G3, R2, W2, Y1) auf. Unter Hypothese 2b tragen letztgenannte Positionen potentiell mit großer Wahrscheinlichkeit zur Spezifität der Allele bei, für die sie einzigartig oder charakteristisch sind.

IV.6 Wie beeinflussen die physikochemischen Differenzen der Aminosäuresubstitutionen und die Aminosäuredeletionen die Ähnlichkeitsdarstellungen der Allele?

Inwiefern wirkt sich die Berücksichtigung der physikochemischen Differenzen auf die einzigartigen, charakteristischen und allgemeinvariablen Merkmale der *csd*-Allele aus? Nach Hypothese 1 sind Allele sich ähnlich, wenn ihre Positionen ähnliche relative und absolute Werte aufweisen. Die Allele weisen dann an den variablen Positionen ähnliche Variabilität auf. Nach Hypothese 2a sind Allele sich ähnlich, wenn ihre Positionen nach Berücksichtigung der physikochemischen Differenzen ähnliche relative Werte haben. Bei Allelen, die nach Hypothese 1 und 2a als ähnlich dargestellt werden, haben die variablen Positionen ähnliche Ränge. Das würde bedeuten, dass bei diesen Allelen auf ähnliche Weise Spezifität kodiert würde, nämlich durch allelinterne Substitutionsmuster. Das besagt, dass das Verhältnis der Aminosäurepositionen eines Allels zueinander für die Spezifität des Allels von Bedeutung

wäre. Je unähnlicher ein Allel zu den anderen Allelen dargestellt wird, desto mehr unterscheiden sich die relativen Werte seiner Positionen von denen der anderen Allele. Besonders einzigartige Positionen verändern die Ränge der Positionen des entsprechenden Allels.

Die Darstellung der Ähnlichkeit der Allele nach Spearman'scher Rangkorrelation mittels ZWHC und MDS ist in den Analysen von Matrix 1, Matrix 2 und Matrix 3 bis auf wenige Unterschiede fast gleich. Es kann also davon ausgegangen werden, dass die relativen Werte der Positionen für die Allele sich durch Berücksichtigung der physikochemischen Differenzen kaum geändert haben. Grob unterteilt werden die Allele B2, B3, G1, G2, R1, R2, R3, Y1 und Y2 als untereinander ähnlich dargestellt und die Allele B1, G3, W1, W2 und Y3 als zu den jeweils anderen 13 Allelen unähnlich. Ein Unterschied zwischen der ZWHC von Matrix 1 und der von Matrix 2 besteht in der Ähnlichkeitsdarstellung der Allele G1, G2 und Y2. Während bei der Analyse von Matrix 1 die Allele G1 und Y2 die größte Korrelation aufweisen, sind es bei der Analyse von Matrix 2 die Allele G2 und Y2. Die Berücksichtigung der physikochemischen Differenz hat offenbar die Rangordnung der Aminosäurepositionen der Allele so beeinflusst, dass die Positionen der Allele G2 und Y2 ähnlichere Ränge haben als die Positionen der Allele G1 und Y2. Wobei die Allele G1 und Y2 von allen Allelen die geringsten Sequenzunterschiede aufweisen, sowohl in Bezug auf die Anzahl der Aminosäuredifferenzen als auch in Bezug auf die Werte der physikochemische Differenzen. Die Positionen der Allele G2 und Y2 haben nach Hypothese 1 also andere relative Werte für die Allele, als nach Hypothese 2a. Nach Hypothese 2a sind daher bei den Allelen G2 und Y2 eine ähnlichere Zusammenstellung von Positionen potentiell für die Spezifität des jeweiligen Allels verantwortlich ist, als nach Hypothese 1.

Ein Unterschied zwischen den MDS nach Spearman'scher Rangkorrelation der Allele von Matrix 1 und Matrix 3 besteht darin, dass die Allele B1 und W1 bei der MDS von Matrix 3 viel näher zusammen liegen, sich also ähnlicher sind, als bei der MDS von Matrix 1. Dies könnte auf die Berücksichtigung der physikochemischen Differenzen zurückzuführen sein. Falls dies der Grund wäre, würden durch Berücksichtigung der physikochemischen Differenzen die Rangordnungen der Positionen dieser beiden Allele so verändert, dass die Positionen potentiell ähnlichere relative Werte für die beiden Allele haben, als ohne Berücksichtigung der physikochemischen Differenzen. Dies würde bedeuten dass nach Hypothese 2a bei den Allelen B1 und W1 eine ähnlichere Zusammenstellung von Positionen potentiell für die

Spezifität des jeweiligen Allels verantwortlich ist, als nach Hypothese 1, wie bei den Allelen G2 und Y2. Durch das Fehlen der Aminosäurepositionen, die Deletionen aufweisen, gehen in Matrix 3 einzigartige oder charakteristische Positionen nicht in die Bewertung ein. Dies betrifft unter anderem die Allele B1 und W1, die viele Deletionen aufweisen und dadurch in Matrix 1 deutlich mehr einzigartige und charakteristische Positionen aufweisen als in Matrix 3. Dies könnte, abgesehen von der Berücksichtigung der physikochemischen Differenzen, der Grund dafür sein, dass bei der MDS von Matrix 3 die Allele B1 und W1 als ähnlicher dargestellt wurden, als bei der MDS von Matrix 1. Bei der ZWHC von M2 ist die größere Ähnlichkeit der Allele B1 und W1 im Vergleich zu der ZWHC von Matrix 1 nachvollziehbar, aber nicht so deutlich erkennbar wie bei den Multidimensionalen Skalierungen.

Nach Hypothese 2b haben die Positionen der Allele, die sich ähnlich sind, nach Berücksichtigung der physikochemischen Differenz ähnliche absolute Werte für die Allele. Je zahlreicher die einzigartigen und charakteristischen Positionen bei einem Allel sind und je größer die physikochemischen Differenzen dieser Positionen sind, desto unähnlicher ist das Allel zu den anderen Allelen. Die Darstellungen der Ähnlichkeiten der Allele zueinander nach Bildung der Euklidischen Distanzen zeigen bei Matrix 1 deutliche Unterschiede zu den Ähnlichkeitsdarstellungen bei den Matrices 2 und 3. Das bedeutet, dass unter Hypothese 2b die Berücksichtigung der physikochemischen Differenzen dazu führte, dass die absoluten Werte der Positionen für die Allele stark verändert wurden im Vergleich zu den absoluten Werten der Positionen ohne Berücksichtigung der physikochemischen Differenzen.

Die Allele B1, B2 und Y3 sind bei der ZWHC von Matrix 2 und der MDS von Matrix 3 nach Bildung der Euklidischen Distanzen den übrigen Allelen deutlich ähnlicher, als bei den Analysen von Matrix 1. Diese Veränderung könnte sowohl auf das Wegfallen der Bewertung von Deletionen als Aminosäureunterschied in den Matrices 2 und 3 zurückzuführen sein als auch auf die Berücksichtigung der physikochemischen Differenzen. Die Allele B1 und B2 besitzen eine oder mehrere einzigartige oder charakteristische Deletionen, so dass bei diesen Allelen die Nichtbewertung der Deletionen als Aminosäuredifferenz vermutlich dazu führte, dass sie den anderen Allelen ähnlicher wurden. Das Allel Y3 besitzt keine einzigartigen oder charakteristischen Deletionen. Bei diesem Allel ist die Steigerung der Ähnlichkeit wahrscheinlich darauf zurückzuführen, dass die physikochemischen Differenzen der einzigartigen oder charakteristischen Aminosäurepositionen dieses Allels meist gering sind. Die Euklidischen Distanzen in den Matrices 2 und 3 verringern sich aus den beiden

genannten Gründen im Vergleich zu den Euklidischen Distanzen in Matrix 1. Nach Hypothese 2b besitzen die Allele B1, B2 und Y3 weniger Positionen, die potentiell für Spezifität verantwortlich sein könnten, als nach Hypothese 1.

IV.7 Aminosäuredeletionen könnten die Spezifität von *csd*-Allelen beeinflussen

Insertionen und Deletionen sind neben den Aminosäuresubstitutionen wichtige Ereignisse in der Evolution von Proteinen (Osuna et al., 2004), die für starke strukturelle Änderungen evolutionär verwandter Proteine verantwortlich sein können (Grishin et al., 2001). Es wurde aber auch berichtet, dass kurze Deletionen oder Insertionen von 1-5 Aminosäuren in Proteinen im Allgemeinen toleriert werden können. Einzelne Deletionen in der Heavy-Chain Complementarity-Determining Region 2 (CDRH2), die für die Antikörperbindung bei humanen Antikörpern zuständig ist, wurden zum Beispiel toleriert, ohne dass die Antikörpererkennung verloren ging (Lantto and Ohlin, 2002).

Die Änderung des Abstandes von Aminosäuren im Protein durch Deletionen kann aber auch dazu beitragen, die Funktion des Proteins zu diversifizieren (Shortle and Sondek, 1995). Das Enzym TEM1 β -Lactamase kann Penicillin hydrolisieren (Osuna et al., 2004). In Spuren ist auch eine Cephalosporin-Hydrolaseaktivität festgestellt worden. Der Omega-Loop des Proteins ist für die Katalyse der Hydrolyse wichtig. Es wurde festgestellt, dass einige Deletionen im Omega –Loop toleriert werden, ohne dass das Enzym seine katalytischen Fähigkeiten vollkommen verliert (Osuna et al., 2004). Die Analyse der Proteinvarianten mit Deletionen im Omega-Loop zeigte, dass die Substratspezifität durch Deletionen von der Penicillinhydrolyse zur Cephalosporinhydrolyse verschoben werden kann (Osuna et al., 2004).

Bei *csd*-Allelen wäre eine Änderung der Spezifität durch Deletionen denkbar. Die potentielle Bedeutung der Positionen, die Deletionen aufweisen (Positionen 144, 211, 307, 358, 359, 362, 372, 373, 417 und 431) für die Spezifität der *csd*-Allele wird im Folgenden analysiert. Die Deletionen an den Positionen 144, 211, 307, 358 und 417 sind jeweils für ein Allel einzigartig. An den Positionen 144, 211 und 307 besteht die Einzigartigkeit durch das Fehlen einer Aminosäure bei jeweils einem Allel (R1, W2 und B2), an den Positionen 358 und 417 durch das Vorhandensein einer Aminosäure bei nur einem Allel (B1). An Position 362 haben die

Allele, die an Position 359 eine Aminosäure haben, eine Deletion und die Allele, die an Position 362 eine Aminosäure tragen haben an Position 359 eine Deletion. Diese beiden Positionen sind charakteristisch für die Allele B1, B3, R3 und W1. Die physikochemischen Differenzen der beiden unterschiedlichen Aminosäuren der Position 359 (Valin und Isoleucin) zu der Aminosäure der Position 362 (Tyrosin) betragen 55 und 33. Falls diese physikochemischen Differenzen keine signifikante Bedeutung für die Proteinstruktur und / oder die Bindungsspezifität haben, könnte dies ein Hinweis darauf sein, dass die beiden Aminosäurepositionen kompensatorisch zusammenwirken. Die Deletionen bei der einen Position würden durch entsprechendes Vorhandensein von Aminosäuren bei der anderen Position kompensiert, so dass sich die Positionen in ihrer Wirkung aufheben und die Tertiärstruktur des Proteins nicht beeinflusst wird. Das Thema Kompensation wird in Abschnitt IV.8 noch einmal betrachtet. Eventuell haben die potentiell kompensatorischen Positionen aber auch Einfluss auf die Spezifität der Allele B1, B3, R3 und W1. Die Positionen 372 und 373 haben beide bei den Allelen R2, W1, W2 und Y1 eine Deletion. Die Allele, die an diesen beiden Positionen keine Deletion aufweisen, tragen an Position 373 alle ein Prolin und an Position 372 vier verschiedene Aminosäuren, nämlich Isoleucin, Valin, Phenylalanin und Leucin. Die physikochemischen Differenzen dieser vier Aminosäuren reichen von 5 bis 50 und sind damit relativ gering. Es könnte sein, dass diese geringen physikochemischen Differenzen wenig oder keine Auswirkungen auf die Proteinstruktur und / oder die Bindungsspezifität der Csd-Proteine haben. Hätten die physikochemischen Differenzen an den Positionen 372 und 373 wirklich keine Bedeutung, so würden die beiden Positionen, die in den ZWHC und MDS als Einzelpositionen dargestellt wurden, zusammen einer Kombination zugeordnet und wären damit potentiell koevolvierte Positionen. Die mögliche Koevolution von Positionen wird in Abschnitt IV.8 diskutiert. Es wäre also denkbar, dass durch das Fehlen (bei den Allelen R2, W1, W2 und Y1) bzw. das Vorhandensein von Aminosäuren (alle übrigen Allele) diese beiden Positionen zusammen die Spezifität von *csd*-Allelen beeinflussen.

IV.8 Es wurden Hinweise auf Koevolution von Aminosäurepositionen gefunden

Der evolutionäre Druck, der auf Proteinen liegt, ist einerseits darauf gerichtet, Aminosäuren zu konservieren, die in die katalytische Aktivität des Proteins involviert sind oder in Interaktionen mit anderen Proteinen (Bashford et al., 1987; Grishin and Phillips, 1994; Ouzounis et al.; 1998; Bartlett et al., 2002), andererseits aber auch darauf Aminosäuren zu konservieren, die an Wechselwirkungen innerhalb des Proteins beteiligt sind (Suel et al., 2003; Zheng et al., 2005) und so dessen Konformation bestimmen. Die Abhängigkeit von Aminosäurepositionen voneinander innerhalb eines Proteins wird als Koevolution bezeichnet (Dutheil and Galtier, 2007). Aminosäurepositionen in einem Protein koevolvieren, wenn der evolutionäre Druck, der aus funktionellen oder strukturellen Gründen auf den Positionen liegt, korreliert ist (Baussand and Carbone, 2009).

Bei der Serin-Protease-Familie zum Beispiel, zu der unter anderem die Verdauungsenzyme Trypsin und Chymotrypsin gehören, sind Aminosäuren, die involviert sind in Proteinfaltung, katalytische Organisation und Ligandenspezifität konserviert innerhalb der Sequenzen der Trypsin und der Chymotrypsin Familien (Baussand and Carbone, 2009). Doch das Ausmaß der Konservierung ist unterschiedlich in Abhängigkeit von der Rolle, die die Aminosäuren spielen. Es wurde herausgefunden, dass die Aminosäuren, die in Proteinfaltung und katalytische Organisation involviert sind, bei allen Serin-Proteasen dieselben sind, während Aminosäuren, die in Ligandenspezifität involviert sind, stark trypsin-oder chymotrypsinspezifisch sind. Die trypsin- und die chymotrypsinspezifischen Aminosäuren standen unter unterschiedlichem evolutionärem Druck, so dass sie familienspezifisch koevolvierten (Baussand and Carbone, 2009). Trypsin und Chymotrypsin zeigen daher familienspezifische Kovariation von Aminosäurepositionen. Aber nicht nur Koevolution kann zur Kovariation von Aminosäurepositionen in evolutionär verwandten Sequenzen führen, sondern auch Hintergrundkovariationen wie der Zufall und die phylogenetische Verwandtschaft (Campo et al., 2008). Wie das Beispiel von Trypsin und Chymotrypsin zeigt, kann auch Bindungsspezifität ein Grund für Kovariation von Aminosäurepositionen sein. Es wäre daher möglich, dass die Spezifität von *csd*-Allelen durch Koevolution von Positionen entstanden ist.

Es könnte untergestellt werden, dass alle variablen Positionen eines *csd*-Allels potentiell koevolviert sind und so gemeinsam zu dessen Spezifität beitragen. Da aber nach Campo und

Mitarbeitern (2008) nicht nur Koevolution dazu führt, dass Positionen in Aminosäuresequenzen Kovariation zeigen, sind vermutlich nicht alle variablen Positionen, an denen sich ein Allel von anderen Allelen unterscheidet, für dessen Spezifität von Bedeutung. In meiner Arbeit wurden bei den Analysen der Matrices 2 und 3 nach Spearman'scher Rangkorrelation Kombinationen von Aminosäurepositionen identifiziert (Abschnitte III.4.1.1.1 und III.4.2.1). Die Positionen einer Kombination haben das gleiche Merkmal: sie sind für die gleichen Allele einzigartig oder charakteristisch oder sie besitzen die gleiche allgemeine Variabilität. Die Positionen jeweils einer Kombination könnten koevolviert sein. Die einzigartigen und charakteristischen Positionen wären dann nur bei den Allelen, bei denen sie verändert sind, koevolviert und die allgemeinvariablen Positionen wären bei allen Allelen koevolviert. Aber auch die Koevolution von Positionen aus verschiedenen Kombinationen und / oder Einzelpositionen wäre denkbar. Als Beispiel sei das Allel B2 genannt. Das Allel besitzt eine Kombination einzigartiger Positionen (bestehend aus drei Positionen) und vier charakteristische Einzelpositionen. Die Spezifität des Allels könnte durch Koevolution dieser sieben Positionen entstanden sein. Eventuell könnten zusätzlich auch allgemeinvariable Positionen mit den einzigartigen und charakteristischen Positionen koevolviert sein, um so Spezifität zu bewirken. Bei der Spezifizierung von *csd*-Allelen kommt es vermutlich nicht nur auf das Merkmal (einzigartig, charakteristisch oder allgemeinvariabel) der Positionen an, sondern vor allem auf die Aminosäuren, die sich an den Positionen befinden. Das würde bedeuten, dass das Zusammenspiel der Aminosäuren, die sich an koevolvierten Positionen befinden, die Spezifität des Allels bestimmt. Die Aminosäurepositionen müssten dabei in der Aminosäuresequenz nicht unbedingt eng zusammen liegen. Studien vieler Proteinkomplexe zeigten, dass Interaktionen von Aminosäuren miteinander, die in der Proteinsequenz weiter auseinander liegen, genauso wichtig für funktionelle Mechanismen des Proteins (Bindungsspezifität, Allosterische Regulation, Konformationsänderungen) sind, wie Aminosäuren, die direkt in Kontakt mit dem Substrat treten (Baussand and Carbone, 2009). Ob Positionen aus verschiedenen Kombinationen und / oder Einzelpositionen bei den verschiedenen Allelen koevolviert sind, kann mit der, in dieser Arbeit angewendeten Analysemethode, nicht ermittelt werden. Doch wie kann festgestellt werden, ob Aminosäurepositionen koevolviert sind und ob es sich bei Kovariation um Hintergrundkovariation handelt? Um die Kovariation, die auf Koevolution beruht von Hintergrundkovariation unterscheiden zu können, wurden viele Methoden

entwickelt (Chothia and Lesk, 1982; Lesk, 1982; Oosawa and Simon, 1986; Altschuh et al., 1988; Bordo and Argos, 1990; Mateu and Fersht, 1999; Lim and Sauer, 1989; Lim et al. 1992; Baldwin et al., 1993; Govindarajan, 2003; Clarke, 1995; Dutheil and Galtier, 2007). Aber die Sensitivität der meisten Methoden ist niedrig. Nach der Methode von Dutheil and Galtier (2007) wird Hintergrundvariation durch Generierung eines phylogenetischen Baumes vor der eigentlichen Analyse der Koevolution deutlich verringert. Die Methode bietet auch die Möglichkeit zu prüfen, ob die Substitutionen kovariierender Positionen sich in ihrer Wirkung ergänzen oder ob sie kompensatorisch (sich in ihrer Wirkung aufhebend) sind.

Aminosäuresubstitutionen, die dazu tendieren, die Struktur des Proteins zu destabilisieren und / oder die Funktion des Proteins zu beeinträchtigen werden eventuell kompensiert durch andere Substitutionen, die dem Protein Stabilität verleihen bzw. die Aufrechterhaltung der Proteinfunktion ermöglichen (Fukami-Kobayashi et al., 2002). Kovariierende Positionen, die sich in ihrer Wirkung ergänzen, könnten in einen adaptiven Prozess involviert sein (Dutheil and Galtier, 2007). Die Selektion in Richtung eines neuen Optimums, das mehr als eine Substitution benötigt, könnte zu Koevolution und damit zu Kovariation von Positionen führen (Or, 2002). Dieses neue Optimum könnte im Falle von Csd-Proteinen die Erlangung neuer, spezifischer Bindungseigenschaften sein, die zu funktionell verschiedenen *csd*-Allelen führen.

Leider ist die Erstellung des phylogenetischen Baumes aus meinen Sequenzdaten statistisch so wenig abgesichert (siehe Abschnitt IV.2), dass ich die Methode von Dutheil and Galtier (2007) nicht zur Analyse verwenden kann. Eine weitere Möglichkeit, Positionen zu identifizieren, die aufgrund von Koevolution kovariieren, geht von der Proteinstruktur aus. Wenn die Proteinstruktur bekannt ist, dann sind die strukturellen und / oder funktionellen Gründe für eine eventuelle Koevolution von Aminosäurepositionen erkennbar, weil ersichtlich ist, welche Positionen in engen Kontakt miteinander treten. Koevolvierende Positionen können so leicht identifiziert werden. Aufbauend auf der Proteinstruktur kann bei Csd aber keine Koevolutionsanalyse von Positionen durchgeführt werden, da die Proteinstruktur größtenteils unbekannt ist.

Wir gehen von der Prämisse aus, dass bei allen Csd-Proteinen die Aminosäurepositionen an der gleichen Stelle sind. Dies setzt voraus, dass die Tertiärstruktur der Proteine trotz der Aminosäuresubstitutionen unverändert bleibt. Bei dem SCR (*S* locus cysteine-rich protein) Gen von *Brassica oleracea* wurde beobachtet, dass die SCR-Proteine trotz starker Variation

ähnliche dreidimensionale Strukturen aufweisen (Chookajorn et al., 2004). Das Aufrechterhalten der Tertiärstruktur trotz Variation könnte durch das Auftreten von kompensierenden Substitutionen gewährleistet werden, die der Veränderung der Tertiärstruktur des Proteins entgegen wirken. Eine physikochemische Differenz, die durch eine Substitution an einer bestimmten Position hervorgerufen wird, kann durch eine oder mehrere Substitutionen an anderen Positionen kompensiert werden (Dutheil and Galtier, 2007). Die kompensierenden Mutationen können in direkter Nachbarschaft zueinander (lokale Kompensationen), aber auch weiter entfernt liegen und erst durch Ausbildung von Sekundärstrukturen in Kontakt mit der zu kompensierenden Mutation treten (Neher, 1994), so dass ein komplexes Netzwerk entsteht. In den von mir identifizierten Kombinationen von Aminosäurepositionen, die für die gleichen Allele einzigartig oder charakteristisch sind oder die gleiche allgemeine Variabilität aufweisen, befinden sich unter anderem auch Positionen, die in direkter Nachbarschaft zueinander liegen. Vermutlich besteht zwischen diesen Positionen eine räumliche Kopplung. Es kann jedoch keine Aussage darüber gemacht werden, ob diese potentiellen Kopplungen kompensatorisch sind oder sich in ihrer Wirkung ergänzen. Substitutionen, die sich in ihrer Wirkung ergänzen würden dafür sprechen, dass durch sie Spezifität entsteht. Aber auch kompensatorische Substitutionen könnten potentiell der Erzeugung von Spezifität dienen, wenn durch sie z.B. Proteinstrukturen aufrecht erhalten werden, die zu spezifischen Bindungseigenschaften von Csd-Proteinen führen.

Koevolvierende Positionen sind häufig in Regionen lokalisiert, die wichtig für die Proteinfunktion sind, wie z.B. aktive Zentren und Oberflächen, die für molekulare Interaktionen mit anderen Proteinen verantwortlich sind (Atchley et al., 2000; Chelvanayagam et al., 1997; Martin et al., 2005; Gloor et al., 2005; Poon and Chao, 2005; Yeang and Haussler, 2007). Die Aminosäurepositionen von elf der zwölf Kombinationen, die ich identifiziert habe, liegen bis auf wenige Ausnahmen in der RS-Domäne und der prolinreichen Region des Csd-Proteins. Nur die Kombination mit den allgemeinvariablen Positionen 7, 23, 24 und 54 liegt im 5' Bereich des *csd*-Gens. Die RS-Domäne und die prolinreiche Region von Csd könnten für die Bindungsspezifität von Csd-Proteinen verantwortlich sein. Die potentiell koevolvierten Positionen, die in der RS-Domäne und der prolinreichen Region liegen, könnten in diesem Fall die Bindungsspezifität der Csd-Proteine beeinflussen. Diese Möglichkeit wird in Abschnitt IV.12 diskutiert.

IV.9 Die hypervariable Region

Die hypervariable Region (HVR) trägt mit großer Wahrscheinlichkeit zur Spezifität von *csd*-Allelen bei (Hasselmann et al., 2008 (1)). Sie liegt innerhalb der PSD von *csd*, zwischen der RS-Domäne und der prolinreichen Region und ist aus Asparagin-Tyrosin (N-Y)-Repeats mit variierenden N-Y Anteilen aufgebaut. Bei den in dieser Arbeit untersuchten *csd*-Allelen enthalten einige der Repeats bei allen Allelen außer den Allelen B3 und G3 noch ein Serin, Cystein, Lysin, Histidin oder Isoleucin.

Die Bedeutung von hypervariablen Regionen für die Spezifität wurde auch schon von anderen Proteinen berichtet. Die hypervariablen Regionen von Immunglobulinen sind ein Beispiel. Deren hypervariable Regionen sind Teil der variablen Domänen und zeigen eine hohe Frequenz von Aminosäuresubstitutionen und große strukturelle Unterschiede (Poljak et al. 1976, Padlan and Davis, 1975; Padlan, 1977). Die funktionale Divergenz der Immunglobuline ist das Ergebnis der strukturellen Divergenz der hypervariablen Regionen, die die Spezifität für das Antigen bestimmt. Dabei können strukturell verschiedene Antikörper die gleiche Bindungsspezifität aufweisen. Dies wurde beobachtet bei Typ II und Typ VIII Antikörpern des Hasen gegen Pneumococccenpolysaccharid (Pincus et al., 1968).

Die hypervariablen Regionen von *csd* sind ebenfalls sehr divers. Sie können aufgrund vieler Deletionen und stark variierender Repeatzahl und -zusammensetzung nicht einem Alignment angeordnet werden und wurden deshalb aus den Zwei-Weg hierarchischen Clusteranalysen und Multidimensionalen Skalierungen ausgeschlossen. Bisher war es noch nicht möglich, die dreidimensionale Struktur von Csd oder auch nur die der hypervariablen Regionen in einem Modell darzustellen. Aufgrund der starken Variationen könnte es sein, dass die verschiedenen hypervariablen Regionen sehr unterschiedliche Tertiärstrukturen aufweisen. Die Aminosäuren, die außer Asparagin und Tyrosin in den Repeats der hypervariablen Regionen der verschiedenen *csd*-Allele zu finden sind, machen die hypervariablen Regionen des jeweiligen Alleles typisch für dieses Allel. Je höher die physikochemischen Differenzen dieser Aminosäuren zu Asparagin und Tyrosin sind, desto größer die Wahrscheinlichkeit, dass diese Aminosäuren zur Spezifität der entsprechenden Allele beitragen. Haben die Aminosäuren also mittlere bis hohe physikochemische Differenzen (>51) ist anzunehmen, dass diese Aminosäuren die Spezifität der jeweiligen Allele beeinflussen können.

Die hypervariablen Regionen der 14 *csd*-Allele habe ich in fünf Typen eingeteilt in Abhängigkeit von der Aminosäurezusammensetzung der Repeats. Die physikochemische Differenz von Asparagin und Tyrosin ist mit 143 hoch. Es ist also vermutlich von großer Bedeutung für die Spezifität von *csd*-Allelen, wie hoch die Asparagin- und Tyrosinanteile an den Repeats sind und aus wie vielen Repeats die HVR besteht.

Typ 1 der hypervariablen Regionen gehören unter anderem die sehr ähnlichen hypervariablen Regionen der Allele G1 und Y2 an. Das Allel G1 besitzt in einem Repeat am 3'-Ende ein Asparagin mehr als das Allel Y2, welches im Vergleich zu dem Allel G1 einen zusätzlichen NNNY-Repeat aufweist. Diese geringen Unterschiede könnten ausreichend dafür sein, Spezifität hervorzurufen. Die Ähnlichkeit der beiden Allele wird in Abschnitt IV.13 näher betrachtet. Das Allel B1 hebt sich durch ein Serin am 3'-Ende, welches eine hohe physikochemische Differenz zu Tyrosin aufweist (146), von den übrigen Typ 1 hypervariablen Regionen ab. Bei den hypervariablen Regionen vom Typ 2 (G3, B3 und Y3) hebt sich das Allel G3 durch eine sehr kurze HVR ab. Die HVR des Allels Y3 enthält ein Lysin innerhalb eines Repeats und ein 3'-endständiges Serin. Lysin hat zu Asparagin und Tyrosin mittlere physikochemische Differenzen von 85 und 94. Bei den hypervariablen Regionen vom Typ 3 (W2, R3, Y1 und R2), fällt besonders die HVR von Y1 auf, die mit einem Cystein endet. Cystein hat zu Asparagin und Tyrosin hohe physikochemische Differenzen von 139 und 194. Die HVR von R2 schließt in ihrem 3'-Bereich ein Lysin in die N-Y-Sequenz ein. Dieses Lysin hebt sich mit physikochemischen Differenzen von 85 zu Tyrosin und 94 zu Asparagin deutlich schwächer ab als das Cystein bei Y1. In der HVR von R3 befinden sich die Aminosäuren Lysin und Serin, die mittlere bzw. hohe physikochemische Differenzen zu Asparagin und Tyrosin aufweisen. Die hypervariablen Regionen der Allele R1 und B2 gehören Typ 4 an und unterscheiden sich lediglich in den letzten beiden Aminosäuren. Ob diese Unterschiede potentiell zur Spezifität der beiden Allele beitragen, kann nicht gesagt werden, da die funktionelle Verschiedenheit der beiden Allele nicht experimentell nachgewiesen wurde. Die HVR des Allels W1 hat sehr wenig mit den hypervariablen Regionen der anderen Allele gemeinsam. Ihr wurde daher ein eigener Typ (5) zugewiesen. Die HVR enthält unter anderem ein Cystein und ein Histidin. Das Cystein besitzt hohe physikochemische Differenzen zu Asparagin und Tyrosin. Das Histidin hebt sich von Asparagin und Tyrosin mit mittleren physikochemischen Differenzen von 68 bis 83 ab.

Die Länge der untersuchten hypervariablen Regionen reicht von sieben Aminosäuren (G3 und W1) bis zu 27 Aminosäuren (Y2). Solche Längendifferenzen führen zu Volumendifferenzen. Es ist wahrscheinlich, dass Volumendifferenzen zu strukturellen Unterschieden der hypervariablen Regionen führen, die auf benachbarte Domänen wie z.B. die RS-Domäne Einfluss haben könnte. Die Struktur und damit die Bindungsspezifität der RS-Domäne könnte beeinflusst werden und damit potentiell auch die Spezifität der Csd-Proteine.

Es kann keine gesicherte Aussage darüber gemacht werden, welche der für die Allele typischen Aminosäuren oder Längendifferenzen der HVR zur Spezifität der *csd*-Allele beitragen und welche nicht. Dies kann nur mit *in vivo* Experimenten (siehe Abschnitt IV.16) geprüft werden. Theoretisch ist eine einzige Aminosäuresubstitution ausreichend, um Spezifität zu erzeugen. Bei S107, einem phosphocholinbindenden Myeloma-Protein wurde beobachtet, dass eine einzige Aminosäuresubstitution in einer hypervariablen, komplementaritätsbestimmenden Region der schweren Kette dazu führte, dass Phosphocholin nicht mehr gebunden werden konnte (Rudikoff et al., 1982). Es wurde dabei eine Glutaminsäure durch ein Alanin ersetzt. Die physikochemische Differenz nach Grantham (1974) ist mit 91 zwar deutlich, gehört aber nicht zu den höchsten feststellbaren physikochemischen Differenzen (für Übersicht über die physikochemischen Differenzen aller Aminosäuren siehe Tabelle II.5). Im dreidimensionalen Modell der Phosphocholin-Bindestelle bildet diese Glutaminsäure an Position 35 eine Wasserstoffbrücke mit einem Tyrosin an Position 94 der leichten Kette. Durch die Substitution wurde die Stabilität der Bindestelle vermutlich beeinträchtigt, was zu veränderten Bindungseigenschaften führte. Aber nicht jede Substitution in der hypervariablen Region führt zu veränderter Bindungsspezifität (Rudikoff et al. 1980). Polymorphismen in den Regionen, die den hypervariablen Regionen von Immunglobulinen benachbart sind, könnten zu der spezifischen Faltung der hypervariablen Domänen beitragen (Mac Kean et al. 1978). Die RS-Domäne und die prolinreiche Region könnten bei Csd-Proteinen Einfluss auf die Faltung der HVR nehmen.

Die Spezifität der Csd-Proteine äußert sich darin, dass zwischen Proteinen mit gleicher Spezifität und Proteinen mit unterschiedlicher Spezifität unterschieden werden kann. Dies ist ein Unterschied zu den Antikörpern, die in der Regel nur ein spezielles Antigen erkennen und binden. Der Mechanismus der spezifischen Erkennung bei Csd-Proteinen unterscheidet sich

daher vermutlich zumindest zum Teil von dem Mechanismus der spezifischen Erkennung bei Antikörpern.

IV.10 Die HVR-Typen sind nicht mit bestimmten Mustern von Aminosäuresubstitutionen in übrigen Teil des *csd*-Gens assoziiert

Sind sich Allele mit ähnlicher HVR auch in den übrigen Regionen des Gens ähnlich?

Ich habe herausgefunden, dass Allele, die mittels der ZWHC und MDS (Abbildungen III.6, III.7, III.8, III.10, III.11, III.13) untersucht und als ähnlich dargestellt wurden sich oft, aber nicht immer, auch in der HVR ähneln, was im Folgenden an Beispielen demonstriert wird. Die Allele G1 und Y2 wurden z.B. in den ZWHC und MDS als ähnlich dargestellt und ihre hypervariablen Regionen wurden demselben Typ zugeordnet. Es gibt aber auch Allele, die in den Zwei-Weg hierarchischen Clusteranalysen und en Multidimensionalen Skalierungen als ähnlich dargestellt wurden, deren hypervariablen Regionen sich aber deutlich unterscheiden. Dies ist zu beobachten bei den Allelen B3 und R3. Es wurden auch Allele gefunden, deren hypervariable Regionen einem Typ angehören, die aber in den Zwei-Weg hierarchischen Clusteranalysen und en Multidimensionalen Skalierungen als nicht sehr ähnlich dargestellt wurden. Dies ist z.B. der Fall bei den Allelen W2 und R2. Es kann daher kein Zusammenhang zwischen der Ähnlichkeit der HVR der *csd*-Allele und der Ähnlichkeit der Allele nach Zwei-Weg hierarchischer Clusteranalyse und Multidimensionaler Skalierung abgeleitet werden. Es ist also davon auszugehen, dass bestimmte HVR-Typen nicht mit bestimmten Mustern von Aminosäuresubstitutionen im anderen Regionen des *csd*-Gens kombiniert sein müssen, damit ein *csd*-Allel Spezifität aufweist.

IV.11 Die Spezifität von *csd*-Allelen ist nicht von einzigartigen Aminosäuresubstitutionen abhängig

Die Ergebnisse der Zwei-Weg hierarchischen Clusteranalysen und der Multidimensionalen Skalierungen zeigen jedoch, dass die Allele B3, G1, R3, Y1 und Y2 keine einzigartigen Positionen besitzen, sondern nur charakteristische und allgemeinvariable. Das bedeutet, dass einzigartige Positionen nicht unbedingt notwendig für die Spezifität der Allele sind. Die Allele G1 und Y2 unterscheiden sich insgesamt nur durch sehr wenige (sechs) Substitutionen voneinander. Dies ist ein Hinweis darauf, dass Spezifität nicht mit der Anzahl der Substitutionen assoziiert ist und dass eine geringe Anzahl an Substitutionen potentiell ausreichend für die Erzeugung von Spezifität ist. Ein Beispiel, bei dem nur wenige Aminosäuren ausreichend sind, um unterschiedliche Spezifität zu erzeugen, sind die SCR-Gene von *Brassica oleracea*. Die SCR-Gene sind über die gesamte Länge sehr divergent, mit nur wenigen konservierten Aminosäuren (Chookajorn et al., 2004). Bei *Brassica oleracea* wurden zwei Allele des SCR-Gens gefunden (SCR₆ und SCR₁₃), zwischen denen nur vier Aminosäurereste darüber entscheiden, ob diese beiden Proteine die gleiche oder unterschiedliche Spezifität aufweisen. Diese vier Aminosäuren (Threonin-Asparagin-Threonin-Glutaminsäure) sind im Strukturmodell des Proteins auf einem oberflächenexponierten Loop lokalisiert und stellen vermutlich den Kontakt zum Rezeptor des SCR-Proteins, der SRK, her (Chookajorn et al., 2004). Es sind hier also wenige Aminosäuresubstitutionen an einer bestimmten Lokalisation im Protein spezifitätsbestimmend.

Die Tatsache, dass die Allele B1, B2, B3, G1, R3 und Y2 nur geringe bis mittlere physikochemische Differenzen aufweisen, könnte ein Hinweis darauf sein, dass physikochemische Differenzen nicht sehr hoch sein müssen, um Spezifität erzeugen zu können. Bei den allgemeinvariablen Aminosäurepositionen 287 und 323 und bei den charakteristischen Aminosäurepositionen 230, 304, 312, 324, 359, 372, 388, 419 und 428 sind mehrere verschiedene physikochemische Differenzen zu beobachten, die teilweise deutliche Unterschiede aufweisen. Position 230, als Beispiel, weist physikochemische Differenzen von 29, 98 und 125 auf. Der Fund von unterschiedlichen physikochemischen Differenzen an Positionen könnte darauf hinweisen, dass die unterschiedlichen physikochemischen Differenzen Bedeutung für die Spezifität der Allele haben könnten. Es

wäre aber auch möglich, dass sich die betrachtete Position an einer Stelle im Protein befindet, an der physikochemische Differenzen keinen oder wenig Einfluss auf die Proteinstruktur und / oder die Spezifität haben, so dass physikochemische Differenzen in einem gewissen Rahmen toleriert werden.

IV.12 Die Spezifität von *csd*-Allelen könnte über spezifische Bindungseigenschaften der RS-Domäne und der prolinreichen Region kodiert werden

Csd ist ein SR-Typ Protein ohne RNA Recognition Motiv (RRM) (Hasselmann et al., 2008 (2)). Es ist bekannt, dass SR-Typ Proteine vermutlich als Moleküle fungieren, die Komplexe im Spleißosom zusammen bringen (Wang et al., 1998; Blencowe et al., 1999). Dies gilt auch für SR-Typ Proteine, die kein RRM besitzen (Cazallah et al., 2005). Die RS-Domänen der SR-Typ Proteine sind involviert in wichtige Protein –Protein Interaktionen (Graveley et al., 1998). Des Weiteren wurde gezeigt, dass RS-Domänen unterschiedliche Bindungsspezifitäten besitzen können (Mattox and Dauwalder, 1998). Proteinbindung wurde auch prolinreichen Regionen zugeschrieben (Kay et al., 2000; Sudol et al., 2001). Wir wissen, dass die RS-Domäne und die prolinreiche Region von Csd proteinbindende Funktion haben (M. Otte, persönliche Mitteilung). Aber sind diese Bindungseigenschaften allelspezifisch?

Wie in Abschnitt IV.8 beschrieben, besteht die Möglichkeit, dass die Spezifität von *csd*-Allelen durch Koevolution mehrerer variabler Aminosäurepositionen entsteht. Koevolvierte Positionen sind häufig in Regionen lokalisiert, die wichtig für die molekularen Interaktionen von Proteinen sind (Atchley et al., 2000; Chelvanayagam et al. 1997; Martin et al. 2005; Gloor et al., 2005; Poon and Chao, 2005; Yeang and Haussler, 2007). Dies würde dafür sprechen, dass vor allem die RS-Domäne und die prolinreiche Region, die 80% der variablen, potentiell koevolvierten Positionen enthalten, für die Bindungsspezifität von Csd-Proteinen verantwortlich sind.

Auch die lokale Umgebung der RS-Domäne hat Einfluss auf die Fähigkeit der Domäne, an Protein-Proteininteraktionen teilzunehmen (Graveley et al., 1998). In unmittelbarer Nähe zur RS-Domäne von Csd liegt die hypervariable Region (HVR). Basierend auf der Annahme von

Graveley und Mitarbeitern (1998) könnte die hypervariable Region, die, wie in Abschnitt IV.9 diskutiert, starke allelische Variation zeigt, Einfluss auf die RS-Domäne und die prolinreiche Region von Csd ausüben und so Bindungsspezifitäten beeinflussen.

In der Einleitung wurden drei hypothetische Modelle zur Proteinassoziation zweier funktionell verschiedener Csd-Proteine vorgestellt (Abschnitt I.3). Bei den Modellen 1 und 2 entsteht der aktivierte Proteinzustand in den Weibchen über unterschiedliche Bindungseigenschaften der funktionell verschiedenen Csd-Proteine. Bei Modell 1 stellen Heterodimere aus zwei funktionell verschiedenen Csd-Proteinen den aktivierten Proteinzustand dar. Bei Modell 2 entstehen keine Heterodimere, sondern Homodimere. Bei Weibchen gibt es dann zwei verschiedene Arten von Homodimeren, bei Männchen nur eine Art. Bei Modell 3 hingegen existieren keine Bindungsunterschiede zwischen funktionell verschiedenen Csd-Proteinen, sondern Aktivitätsunterschiede. Die Modelle 1 und 2 wären durch spezifische Bindungseigenschaften der RS-Domäne und der prolinreichen Region zu erklären. Die Koevolution von Aminosäuren bestimmter Positionen der RS-Domäne und der prolinreichen Region bei den verschiedenen *csd*-Allelen könnte dazu beitragen, die Bindungseigenschaften dieser Domänen allelspezifisch zu verändern. Das bedeutet, dass die Anzahl der Aminosäuresubstitutionen und -deletionen und die physikochemischen Eigenschaften der substituierten Aminosäuren in der RS-Domäne und der prolinreichen Region wichtig für die spezifischen Bindungseigenschaften von Csd-Proteinen sein könnten. Auch die unterschiedliche Verteilung der Aminosäuresubstitutionen und -deletionen in der RS-Domäne und der prolinreichen Region könnte die Bindungsspezifität der Csd-Proteine beeinflussen. Vielleicht ist es z.B. von Bedeutung, ob die Substitutionen eher auf den vorderen Bereich der RS-Domäne konzentriert sind, wie bei dem Allel W2 oder auf den hinteren Bereich, wie bei dem Allel R1, oder darauf, dass in der prolinreichen Region mehr Substitutionen stattfanden als in der RS-Domäne, wie bei dem Allel Y1. Auch die Verteilung der Deletionen in der prolinreichen Region könnte starken Einfluss auf die Bindungsspezifität der Allele nehmen. Außerdem könnte die HVR des *csd*-Gens entscheidend zur Proteinbindungsspezifität von Csd-Proteinen beitragen. Entweder direkt, durch allelspezifische Aminosäuresubstitutionen und -deletionen, die zu spezifischer räumlicher Struktur und damit zu spezifischer Bindungsaffinität der HVR führen könnten, oder indirekt durch Beeinflussung der Bindungseigenschaften der RS-Domäne. Nach Modell 1 würden dann die potentiell für Bindungsspezifität verantwortlichen Aminosäurepositionen in der RS-

Domäne und der prolinreichen Region dafür sorgen, dass sich nur funktionell unterschiedliche Csd-Proteine aneinander binden können, um Heterodimere zu bilden. Das aktive Protein kann dann über Bindung an weitere Proteine des Spleißosoms den Spleißvorgang von *fem* pre-mRNA über die Wahl der Spleißstellen so beeinflussen, dass weibchenspezifisches Transkript entsteht.

Nach Model 2 würden die spezifischen Bindungseigenschaften dafür sorgen, dass nur Csd-Proteine mit derselben Bindungsspezifität aneinander binden können. Nach diesem Modell wäre es möglich, dass das Spleißosom nur durch zwei verschiedene Homodimere so vervollständigt werden kann, dass diejenigen Spleißstellen gewählt werden, die zu einem weibchenspezifischen Transkript von *fem* führen. In diesem Fall wäre dann kein System erforderlich, welches den Dosisunterschied der Homomere misst, wie zuerst angenommen.

Die Spezifität der Csd-Proteine könnte auch auf andere Besonderheiten als die Bindungsspezifität funktionell verschiedener Csd-Proteine zurückzuführen sein. Das Modell der funktionellen Komplementation (Modell 3) sieht keine Bindungsunterschiede zwischen funktionell verschiedenen Csd-Proteinen vor, sondern Aktivitätsunterschiede. Es kann daher nicht über die Bindungsspezifität der RS-Domäne erklärt werden. Nach diesem Modell liegen die Csd-Proteine in aktivem Zustand als zwei verschiedene Monomere oder als Heterodimer vor, die sich durch Aktivitätsunterschiede funktionell komplementieren. Die Aminosäuresubstitutionen zwischen den Allelen könnten potentiell dazu führen, dass funktionell verschiedene Csd-Proteine unterschiedliche Konformationen aufweisen, die dann zu unterschiedlichen Aktivitätszuständen führen würden. Wir gehen allerdings davon aus, dass keine Konformationsunterschiede zwischen den funktionell verschiedenen Csd-Proteinen bestehen.

IV.13 Die Allele G1, G2 und Y2 zeigen Spuren eines Rekombinationsereignisses

Die Allele G1, G2 und Y2 wurden für eine separate Untersuchung herausgegriffen, da sie in allen Zwei-Weg hierarchischen Clusteranalysen und Multidimensionalen Skalierungen als

einander ähnlich dargestellt wurden, aber nachweislich funktionell verschieden sind. Die Ursache für die funktionelle Verschiedenheit dieser Allele kann daher auf wenige Aminosäurepositionen eingeschränkt werden. Die Allele G1, G2 und Y2 haben außerdem hypervariable Regionen des gleichen Typs. Die Rekonstruktion eines Baumes und der Test auf Phylogenie zeigten jedoch, dass von den 14 *csd*-Allelen nur die Allele G1 und Y2 mit großer Wahrscheinlichkeit phylogenetisch eng verwandt sind. Das bedeutet, dass diese beiden Allele vermutlich erst kürzlich voneinander divergiert sind. Aminosäureunterschiede zwischen diesen beiden Allelen sind daher mit hoher Wahrscheinlichkeit von Bedeutung für die Spezifität dieser Allele. Die Aminosäuresequenzen der Allele G1 und Y2 unterscheiden sich außerhalb der HVR nur an den sechs Positionen 7, 23, 24, 42, 54 und 93. Die physikochemischen Differenzen dieser Positionen liegen zwischen 46 und 64 und sind damit nicht sehr ausgeprägt. Ob diese Positionen dennoch zur Spezifität von G1 und Y2 beitragen, kann nicht bestimmt werden. Es wäre möglich, dass die Positionen der Erhaltung der Tertiärstruktur der Csd-Proteine dienen, oder aber nicht durch negative Selektion entfernt wurden, da sie die Funktionsfähigkeit von Csd nicht beeinflussen. In diesem Fall wäre nur der Unterschied der hypervariablen Regionen von G1 und Y2 entscheidend für die Spezifität der Allele. Die hypervariablen Regionen dieser beiden Allele sind sich sehr ähnlich. Sie unterscheiden sich neben einem geringfügigen Unterschied der N-Y-Zusammensetzung eines Repeats im 3' Bereich nur durch eine Längendifferenz von drei Aminosäuren.

Die Allele G2 und Y2 unterscheiden sich an 16 Aminosäurepositionen außerhalb der HVR, wobei die Positionen 219 und 293 aufgrund hoher physikochemischer Differenzen (94 und 125) mit größerer Wahrscheinlichkeit zu Spezifität beitragen, als die anderen variablen Positionen. Die hypervariablen Regionen der beiden Allele unterscheiden sich durch eine Längendifferenz von 13 Aminosäuren und einen geringfügigen Unterschied der N-Y-Zusammensetzung eines Repeats. Sie sind sich daher nicht so ähnlich, wie die hypervariablen Regionen der Allele G1 und Y2.

Wie in Abbildung III.19 zu sehen ist, unterscheidet sich das Allel G1 nur im 5'-Bereich des Gens von dem Allel Y2, während sich das Allel G2 vom mittleren Bereich des Gens bis zum 3'-Ende von dem Allel Y2 unterscheidet. Die beiden G- Allele haben keine gemeinsamen Aminosäuredifferenzen zu dem Allel Y2. Die elf anderen Allele des Kreuzungsexperimentes weisen Aminosäureunterschiede zu diesen drei Allelen auf, die über das ganze *csd*-Gen verteilt sind. Gemeinsame Cluster polymorpher Positionen könnten ein Hinweis auf das

Vorkommen von Genkonversion sein, die durch Rekombination entsteht (Hasselmann and Beye 2006). Die Ähnlichkeit der drei Allele G1, G2 und Y2 könnte daher folgendermaßen erklärt werden: Zwischen einem Vorläuferallel der beiden G-Allele (GX) und dem Allel Y2 fand eine Rekombination statt (Abbildung IV.1.A). Die entstandenen Allele G1 und G2 trugen nun jeweils eine Hälfte des GX Allels und eine Hälfte des Y2 Allels. Es wäre aber auch möglich, dass das Allel Y2 durch Rekombination der Allele G1 und G2 entstand (Abbildung IV.1.B). Nach beiden Hypothesen müssten die hypervariablen Regionen der Allele G1 und Y2 sich ähnlicher sein als die der Allele G2 und Y2. Dies ist auch der Fall.

Wären die Aminosäuresubstitutionen außerhalb der PSD von *csd* bedeutend für Spezifität, so wäre hiermit gezeigt, dass diese Spezifität nicht auf eine bestimmte Genregion festgelegt ist. Denn bei den Allelen G1 und Y2 unterscheidet sich die 5' Region des *csd*-Gens und die Allelen G2 und Y2 zeigen von Position 190 bis zum 3' Ende Aminosäureunterschiede.

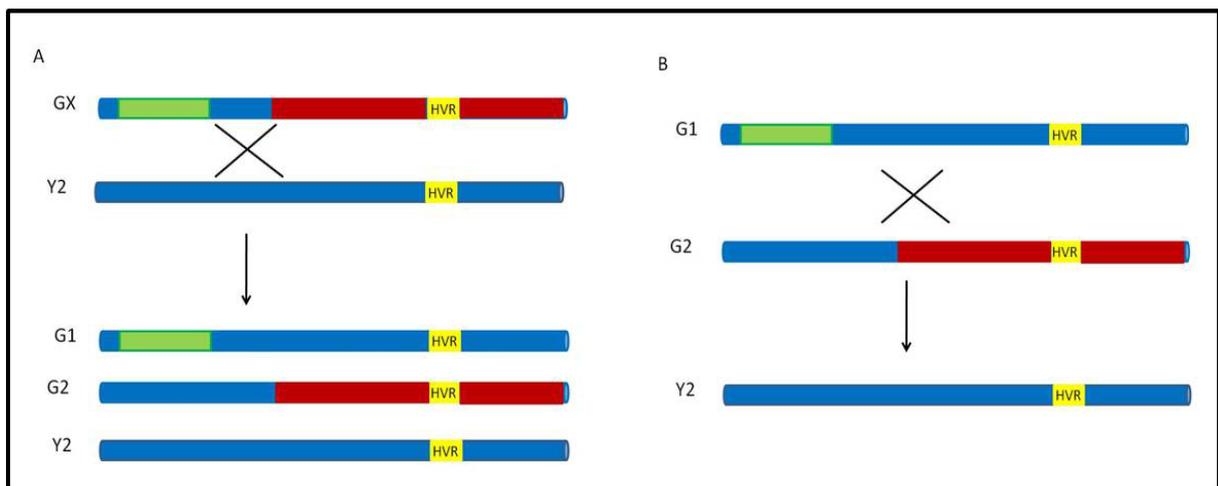


Abbildung IV.1: A: Die Allele G1 und G2 könnten durch Rekombination eines Vorläuferallels GX mit dem Allel Y2 entstanden sein. B: Das Allel Y2 könnte durch Rekombination aus den Allelen G1 und G2 entstanden sein. Die Regionen an denen das Allel G1 Aminosäureunterschiede zu dem Allel Y2 aufweist sind grün markiert, die Regionen, in denen das Allel G2 zu dem Allel Y2 Aminosäureunterschiede aufweist, sind rot. Die hypervariable Region ist unabhängig von allelelischer Variabilität gelb dargestellt.

IV.14 Diskussion der Methoden

Bei den beiden Analysemethoden Zwei-Weg hierarchischen Clusteranalyse (ZWHC) und Multidimensionale Skalierung (MDS) werden die Ergebnisse graphisch auf sehr unterschiedliche Weise dargestellt. Beide Darstellungsweisen haben Vor- und Nachteile.

Bei der ZWHC werden die Ähnlichkeiten der Allele und Aminosäurepositionen parallel in einer Abbildung dargestellt. Es ist an der Graphik erkennbar, wie hoch die Positionswerte jeder Aminosäureposition bei jedem Allel sind. Gemeinsamkeiten und Unähnlichkeiten zwischen Allelen und zwischen Aminosäurepositionen können erkannt werden. Aus der Abbildung kann daher entnommen werden, welche Positionswerte dafür verantwortlich sind, dass Allele oder Aminosäurepositionen als ähnlich oder unähnlich dargestellt werden. Es ist zum Beispiel gut zu erkennen, ob sich ähnelnde Positionen einzigartig, charakteristisch oder allgemeinvariabel sind. Es kann ein Zusammenhang hergestellt werden zwischen der Ähnlichkeit der Allele und der Aminosäurepositionen. Daraus ist erkennbar, welche Positionen für die Spezifität von Allelen potentiell mehr bzw. weniger Bedeutung haben. Weiterhin ist ersichtlich, nach welchen Kriterien die Allele oder die Positionen als ähnlich oder unähnlich dargestellt werden. Es ist z.B. zu sehen, ob Aminosäurepositionen nach den Merkmalen, die sie aufweisen oder vorrangig nach der Höhe ihrer Positionswerte als ähnlich dargestellt werden. An der Graphik ist auch erkennbar, wenn ein Positionswert evtl. nur beschränkt aussagekräftig ist. Der Positionswert ist dann ganz schwarz (Positionswert beträgt null) weil bei Erstellung der Matrix das betrachtete Allel nicht mit dem Allel verglichen wurde, das an der betrachteten Position eine einzigartige Aminosäure aufweist. Außerdem ist erkennbar, welche Positionswerte aufgrund von Deletionen gar nicht in die Analyse eingegangen sind. Ein großer Nachteil der ZWHC ist der Verlust von Information über die Ähnlichkeit der Allele und Positionen durch Dichotomie, die durch die Hierarchie des Clusterverfahrens hervorgerufen wird.

Bei den Ähnlichkeitsdarstellungen, die mittels MDS durchgeführt wurden, besteht der Nachteil der Dichotomie nicht. Die Ähnlichkeiten der Allele bzw. der Positionen sind durch die Daten der (Un)Ähnlichkeitsmatrices gegeben und werden in der Ordination möglichst entsprechend der Abstände wiedergegeben, die die Daten in der (Un)Ähnlichkeitsmatrix haben. Die Ähnlichkeit aller Allele bzw. Aminosäurepositionen zueinander wird daher durch die Abstände der Datenpunkte, die die Allele bzw. die Positionen darstellen, ausgedrückt. Bei

guten Stresswerten sollte die Darstellung einer MDS sehr genau sein. Ähnlichkeiten und Unähnlichkeiten sind besser erkennbar, als bei der ZWHC, da der Nachteil der Dichotomie nicht besteht. Allele mit vielen einzigartigen Aminosäurepositionen und Aminosäurepositionen mit hohen physikochemischen Differenzen liegen zum Beispiel deutlich separiert von den anderen Allelen bzw. Aminosäurepositionen in den Randgebieten der Graphiken. So können Änderungen der Ähnlichkeit von Allelen oder Aminosäurepositionen, die zum Beispiel durch Berücksichtigung der physikochemischen Differenzen entstanden, mittels Vergleich der Ordinationen von Matrix 1 und Matrix 3 leichter erkannt werden als durch Vergleich der Clusteranalysen der Matrices 1 und 2. Bei den ZWHC von Matrix 2 ist zum Beispiel die größere Ähnlichkeit der Allele B1 und W1 im Vergleich zu der ZWHC von Matrix 1 nachvollziehbar, aber nicht so deutlich erkennbar wie bei den Ordinationen von Matrix 1 und Matrix 3. Ein großer Nachteil der MDS gegenüber der ZWHC ist, dass die Ähnlichkeit von Allelen und Aminosäurepositionen nicht parallel dargestellt werden kann. Aus der Graphik gehen keine Informationen über die Höhe der Positionswerte hervor, so dass kein Zusammenhang zwischen der Ähnlichkeit der Allele und der Aminosäurepositionen hergestellt werden kann. Es ist zwar sichtbar, welche Allele oder Aminosäurepositionen sich ähneln, aber nicht wodurch Ähnlichkeit oder Unähnlichkeit hervorgerufen wird. Bei umfangreichen Datensätzen, wie zum Beispiel den der Aminosäurepositionen wird die Ordination sehr unübersichtlich und es müssen Koordinatenwerte zu Hilfe genommen werden, um Positionen identifizieren zu können. Die Stresswerte sind bei den großen Datensätzen der Positionen zum Teil so hoch, dass die Ordination allenfalls zur groben Abschätzung der Ähnlichkeiten von Positionen dienen kann. Und selbst bei kleineren Datensätzen, wie denen der Allele, bei denen die Stresswerte der Ordinationen eine Höhe haben, die auf eine noch einigermaßen zuverlässige Darstellung der Ähnlichkeiten schließen lässt, werden Ähnlichkeiten nicht korrekt dargestellt. Die Allele G1 und Y2 weisen zum Beispiel bei der Spearman'schen Rangkorrelation von Matrix 1 den größten Korrelationskoeffizienten auf (Tabelle A.XV) und bei Bildung der Euklidischen Distanzen die geringste Distanz (Tabelle A.XVI). Sie müssten also in den Ordinationen der Allele von Matrix 1 den geringsten Abstand haben, was nicht der Fall ist. Vermutlich geben die Ordinationen mit Stresswerten von 0,18 (Matrix 1) und 0,17 (Matrix 3) die Abstände der Datenpunkte, die die Allele darstellen, nicht so wieder, wie sich die Ähnlichkeiten der Allele aus den Ähnlichkeitsmatrices ergeben.

Die ZWHC hat zwar, wie dargestellt, auch Nachteile, ist für meine Analysen aber viel besser geeignet, als die MDS. Anhand einer ZWHC Abbildung kann ein Überblick über Ähnlichkeiten der Allele und der Aminosäurepositionen gewonnen werden und gleichzeitig erkannt werden, worin diese Ähnlichkeiten bestehen. Der Informationsgehalt der MDS ist deutlich geringer als der der ZWHC. Der einzige Vorteil der MDS im Vergleich zur ZWHC ist die objektivere Darstellung der Ähnlichkeit der Daten. Dieser Vorteil wird allerdings dadurch aufgehoben, dass die Ordinationen die Ähnlichkeiten nicht zuverlässig korrekt wiedergeben, trotz akzeptabler Stresswerte.

IV.15 Fehlerdiskussion

Die Art der Matrixbildung birgt Quellen für Ungenauigkeiten, die bei den ZWHC und MDS eventuell zu ungenauen Ergebnissen führten. Es wurde experimentell nicht jedes Allel mit jedem kombiniert. Die verschiedenen Allele waren daher in unterschiedlich viele Vergleiche involviert. Daraus folgt, dass die Positionswerte der einzelnen Allele nicht aus der gleichen Anzahl von Werten berechnet wurden und daher nicht den gleichen Informationsgehalt haben. Da Positionswerte Mittelwerte sind, sind sie abhängig von der Zahl der Allele, die in die Berechnung eingehen und den Aminosäuren, die die Allele an der betrachteten Position tragen. Dies führt dazu, dass z.B. Positionswerte zweier Allele an einer betrachteten Aminosäureposition unterschiedlich hoch sein können, auch wenn die beiden Allele die gleiche Aminosäure besitzen.

Bei Erstellung der (Un)Ähnlichkeitsmatrices werden die Positionswerte der Allele oder der Aminosäurepositionen paarweise verglichen und deren Ähnlichkeit mittels des Rangkorrelationskoeffizienten oder der Euklidischen Distanz in der (Un)Ähnlichkeitsmatrix angegeben. Die angegebene (Un)Ähnlichkeit sagt nichts darüber aus, ob sich die Allele oder Positionen wirklich unterscheiden bzw. gleich sind, da die Positionswerte nur die durchschnittliche Variabilität eines Allels an einer Aminosäureposition angeben. In den (Un)Ähnlichkeitsmatrices ist daher nicht die Information enthalten, ob zum Beispiel zwei verglichene Allele an ihren Aminosäurepositionen wirklich die gleiche Aminosäure aufweisen, oder ob sie nur die gleiche Variabilität aufweisen.

Die Darstellungen der Ähnlichkeiten von Allelen und von Aminosäurepositionen mit Hilfe von ZWHC und MDS könnten durch Fehlen von Information in den Matrices und (Un)Ähnlichkeitsmatrices ungenau oder fehlerhaft sein. Dies ist bei dieser Art der Datenanalyse des gegebenen Kreuzungsexperimentes nicht zu vermeiden.

Eine weitere, durch den experimentellen Aufbau bedingte, Fehlerquelle ist bei der Rangvergabe zu finden, die vor der Spearman'schen Rangkorrelation vorgenommen wird. Die Wahrscheinlichkeit ist hoch, dass durch geringe Unterschiede in den Positionswerten, die experimentell bedingt sind und keine Bedeutung haben, die Ränge falsch vergeben werden. Experimentell bedingte Unterschiede können entstehen, weil nicht jedes Allel mit jedem verglichen wurde. Wenn zum Beispiel ein Allel an einer Position zu allen anderen Allelen eine Differenz aufweist, und von den übrigen Allelen manche experimentell mit diesem verglichen wurden, andere aber nicht, so haben die experimentell paarweise verglichenen Allele an dieser Position einen anderen Positionswert, als die nicht verglichenen Allele. Die Positionswerte sind unterschiedlich, obwohl alle, bis auf das eine Allel, die gleiche Aminosäure tragen. Die Unterschiede an dieser Position dürften daher nicht als Unterschiede bewertet werden. Dies wirkt sich besonders auf die Darstellung der Allelähnlichkeit aus. Da jedes Allel aus 82 (Matrices 1 und 2) bzw. 72 (Matrix 3) Aminosäurepositionen besteht, werden hier entsprechend viele Ränge vergeben. Da bei der Rangbildung der Aminosäurepositionen nur 14 Ränge (entsprechend der 14 Allele) vergeben werden, fällt der Fehler durch Rangbildung vermutlich nicht so groß aus wie bei den Allelen. Verbessern würde sich diese Situation, wenn der Stichprobenumfang stark erhöht würde, also mehr Allele in die Kreuzung eingebracht würden, so dass es nicht ins Gewicht fiel, wenn manche paarweisen Vergleiche nicht existieren. Der Fehler könnte ganz vermieden werden, wenn jedes, ins Kreuzungsexperiment eingebrachte Allel mit jedem paarweise verglichen würde.

Die physikochemischen Differenzen, die nach Grantham's Methode (1974) ermittelt wurden, sind eine gute Näherung, eine perfekte Wiedergabe des physikochemischen Unterschiedes zwischen Aminosäuren sind sie jedoch nicht. Dies stellt eine Fehlerquelle dar, die in meiner Arbeit zu Fehleinschätzungen der Unterschiede zwischen Allelen oder Aminosäurepositionen geführt haben könnte.

IV.16 *In vivo* Versuche zum Testen bestimmter Genregionen von *csd* auf ihre Bedeutung für die Spezifität von *csd*-Allelen

Um bestimmte Genregionen von *csd* darauf zu testen, ob sie Einfluss auf die Spezifität von Allelen haben, werden Aminosäurepositionen betrachtet, die bei Paaren nachweislich funktionell verschiedener Allele variabel sind. Die funktionell verschiedenen Allele G1 und Y2 eignen sich hierfür, denn sie unterscheiden sich sowohl außerhalb der HVR als auch in der HVR nur geringfügig, so dass die Zahl der potentiell für die Spezifität wichtigen Aminosäuren auf sechs Positionen im 5'-Bereich des *csd*-Gens und die HVR selber eingegrenzt werden kann.

Es werden zurzeit Spleißassays von Dipl.-Biol. Inga Nissen etabliert, die es ermöglichen sollen, in Zellkultur (Schmetterlingszellen) zwei *csd*-Allele zu exprimieren und anhand des Spleißmusters von *fem* die funktionelle Verschiedenheit dieser Allele zu testen. Sobald diese Assays etabliert sind, könnten folgende funktionelle Tests durchgeführt werden:

1. Um zu testen, ob die HVR alleine Spezifität hervorrufen kann, werden zwei *csd*-Allele synthetisiert, die bis auf die HVR entweder die Aminosäuresequenz des Allels G1 oder des Allels Y2 haben. Eines der synthetisierten Allele trägt die HVR des Allels G1, das andere die des Allels Y2. Die Allele werden gleichzeitig in den Zellen exprimiert und wie angegeben getestet, ob die Allele funktionell verschieden sind.
2. Um zu testen, ob Aminosäureunterschiede außerhalb der HVR ausreichend sind, Spezifität hervorzurufen, werden zwei *csd*-Allele synthetisiert, die außerhalb der HVR die Aminosäuresequenzen der Allele G1 und Y2 haben. Beide synthetisierten Allele tragen entweder die HVR des Allels G1 oder die des Allels Y2.
3. Konnte die Spezifität ausschließlich der Region außerhalb der HVR oder ausschließlich der HVR zugeordnet werden, muss noch ermittelt werden, welche Aminosäurepositionen genau die Spezifität hervorrufen. Dafür werden die Region, die nicht für Spezifität wichtig ist und die Region, die Spezifität bewirkt, bei beiden synthetischen Allelen zunächst konstant gehalten (entweder die Sequenz des Allels G1 oder die des Allels Y2). Schrittweise werden nun die Aminosäuresubstitutionen, die zwischen den Allelen G1 und Y2 gefunden wurden, bei einem der beiden synthetischen Allele einzeln eingefügt und jeweils der Einfluss auf die Spezifität überprüft.

Kann bei den einzeln eingefügten Substitutionen keine Spezifität festgestellt werden, so ist es wahrscheinlich, dass mehrere Aminosäuresubstitutionen zusammen die Spezifität

kodieren, so dass dann auch verschiedene Aminosäurepositionen in unterschiedlicher Zusammensetzung substituiert werden müssen.

Kann Spezifität weder der Region außerhalb der HVR noch der HVR zugeordnet werden, ist es wahrscheinlich, dass beide Regionen zur Spezifität beitragen. In diesem Fall müssten schrittweise alle möglichen Aminosäurepositionen in unterschiedlicher Zusammensetzung innerhalb und außerhalb der HVR substituiert werden.

Für den Fall, dass auf Zellkulturebene keine Nachweise durchgeführt werden können, gibt es die Möglichkeit von Experimenten an Bienenembryonen. Da es nur schwer möglich ist, in der Natur die Allele G1 und Y2 zu finden, müssen diese Tests mit anderen Allelen durchgeführt werden. Bei solchen Experimenten wird dann in ein haploides oder homozygotes Ei, bei dem die Aminosäuresequenz des *csd*-Allels ermittelt wurde, mRNA eines *in vitro* synthetisierten *csd*-Allels injiziert. Dieses synthetisierte Allel entspricht dem Allel des Eis, in das aber Substitutionen eingefügt wurden. Die Region der PSD, die sich in unmittelbarer Nachbarschaft der HVR befindet, trägt mit großer Wahrscheinlichkeit zur Kodierung von Spezifität bei (Hasselmann et al. 2008 (1)). Aminosäuresubstitutionen in dieser Region sollten daher zuerst zusammen mit Aminosäuresubstitutionen in der HVR getestet werden, bevor auch Aminosäuresubstitutionen im 5'-Bereich des *csd*-Gens untersucht werden. Im 5'-Bereich des Gens könnten dann die als allgemeinvariabel identifizierten Positionen getestet werden, die bei den Allelen G1 und Y2 verschieden waren.

Bei der HVR könnten die Effekte von Veränderungen in der Zusammensetzung und der Anzahl der N-Y-Repeats auf die Spezifität analysiert werden durch das Einfügen von Aminosäuren, die große oder nur geringe physikochemische Differenzen verursachen. Auch die Länge der HVRs könnte variiert werden.

Zusammenfassung

Die Geschlechtsbestimmung bei der Honigbiene *Apis mellifera* erfolgt über den allelischen Zustand des *csd*-Gens. Individuen, die am *csd*-Lokus heterozygot sind, also zwei spezifische (funktionell verschiedene) *csd*-Allele tragen, entwickeln sich zu Weibchen, am *csd*-Lokus homo- oder hemizygot Individuen zu Männchen. Homozygote Männchen werden kurz nach dem Schlupf von Arbeiterinnen gefressen, so dass nur heterozygote und hemizygot Individuen das Imaginalstadium erreichen und sich zu Arbeiterinnen bzw. Königinnen und Drohnen entwickeln. Um herauszufinden, welche Unterschiede in der Aminosäuresequenz die Spezifität von *csd*-Allelen kodieren, wurden die paarweisen Aminosäurepolymorphismen zwischen den funktionell verschiedenen *csd*-Allelen von heterozygoten Arbeiterinnen analysiert. Insgesamt wurden die Aminosäurepolymorphismen von 14 *csd*-Allelen in 71 verschiedenen paarweisen Zusammenstellungen untersucht. Mit Hilfe von Zwei-Weg hierarchischer Clusteranalyse (ZWHC) und Multidimensionaler Skalierung (MDS) wurden die variablen Aminosäurepositionen zwischen den Allelen, die außerhalb der hypervariablen Region des *csd*-Gens liegen, analysiert und in Abhängigkeit folgender Merkmale in drei Kategorien unterteilt: 1. einzigartige Positionen, die nur bei einem *csd*-Allel verändert sind, 2. charakteristische Positionen, die bei zwei bis vier Allelen verändert sind und 3. allgemeinvariable Positionen, die bei allen Allelen Variabilität aufweisen. Bei fünf von 14 Allelen wurden keine einzigartigen Positionen gefunden. Einzigartige Aminosäuresubstitutionen scheinen daher nicht notwendig für die Spezifität von *csd*-Allelen zu sein. Die Analysen zeigten außerdem, dass es funktionell verschiedene *csd*-Allele gibt, die sich nur durch sehr wenige (sechs) Aminosäuresubstitutionen unterscheiden. Dies ist ein Hinweis darauf, dass Spezifität durch wenige Substitutionen kodiert werden kann. Werden die Änderungen der physikochemischen Eigenschaften, die bei den Aminosäuresubstitutionen entstehen, bei den Analysen mittels ZWHC und MDS berücksichtigt, zeigt sich, dass Substitutionen, die geringe bis mittlere Änderungen der physikochemischen Eigenschaften nach sich ziehen, potentiell ausreichend dafür sind, Spezifität zu induzieren. Die Tatsache, dass 80% der variablen Aminosäurepositionen, die bei den untersuchten *csd*-Allelen gefunden wurden, in der RS-Domäne und der prolinreichen Region des *csd*-Gens lokalisiert sind, macht es wahrscheinlich, dass Spezifität zumindest zum Teil über spezifische Bindungseigenschaften der RS-Domäne und der prolinreichen Region von Csd-Proteinen erzielt wird. Die hypervariable Region des *csd*-Gens zeigt starke allelische Variabilität und könnte ebenfalls dazu beitragen, die Spezifität von *csd*-Allelen zu kodieren. Sie besteht aus Asparagin-Tyrosin (N-Y)-Repeats. Die Repeats sind bei verschiedenen *csd*-Allelen in unterschiedlicher Anzahl vorhanden und weisen variierende N-Y Anteile auf. Bei den meisten Allelen befinden sich in einem oder zwei der Repeats außerdem die Aminosäuren Serin, Cystein, Lysin, Histidin oder Isoleucin. In Abhängigkeit von der Aminosäurezusammensetzung der Repeats wurden die hypervariablen Regionen der 14 *csd*-Allele in fünf Typen unterteilt. In den Analysen konnte kein Zusammenhang zwischen der Typzugehörigkeit der hypervariablen Region und der Ähnlichkeit der Allele nach ZWHC und MDS festgestellt werden. Daraus kann geschlossen werden, dass HVR-Typen nicht mit

bestimmten Mustern von Aminosäuresubstitutionen in übrigen Teil des *csd*-Gens kombiniert sein müssen, damit ein *csd*-Allel Spezifität aufweist.

Abstract

In the case of the honeybee *Apis mellifera*, sex is determined by the allelic condition of the *csd* gene. Individuals, that are heterozygous at the *csd* locus have two specific (functional different) *csd* alleles. Heterozygous individuals develop into females whereas individuals, which are homo- or hemizygous at *csd* develop into males. Homozygous males are eaten by workerbees shortly after they hatch from egg. Only hetero- and hemizygous individuals grow up to become adult workerbees and drones. To identify the amino acid differences encoding the specificity of *csd* alleles, the pairwise amino acid polymorphisms between the functional different *csd* alleles of heterozygous workerbees were compared. Altogether the amino acid polymorphisms of 14 *csd* alleles in 71 different pairwise combinations were examined.

The amino acid positions that are outside the hypervariable region of the *csd* gene and that are variable between *csd* alleles were analysed by two-way hierarchical clustering and multidimensional scaling. The variable Positions found have been subdivided into unique, characteristic and generally variable Positions. Unique positions are changed only at one allele, characteristic positions are changed at two up to four alleles and generally variable positions are variable at all alleles. Five of the 14 *csd* alleles showed no unique positions. Therefore, unique Positions do not seem to be necessary for specificity. Moreover the two-way hierarchical clustering and the multidimensional scaling showed that some functional different alleles only have few (six) amino acid differences. This indicates that specificity can be encoded by a small number of substitutions. When amino acids are substituted, the physicochemical properties change. If these changes are taken into account when doing two-way hierarchical Clustering and the multidimensional scaling it turns out, that small to medium changes in physicochemical properties are potentially sufficient to induce specificity.

80% of the variable amino acid positions of the examined *csd* alleles have been found in the RS-domain and the proline-rich region of the *csd* gene. Due to this fact it is probable, that specificity is encoded at least to some extent, by specific binding properties of the RS-domain and the proline-rich region from Csd proteins.

The hypervariable region of the *csd* gene shows strong allelic variability. Therefore, this region could contribute to specificity of *csd* alleles as well. The hypervariable region consists of asparagines-tyrosine (N-Y) repeats. These repeats exist in different numbers and consist of variable compositions of asparagine and tyrosine in different *csd* alleles. Besides this, most of the alleles show the amino acids cysteine, lysine, histidine oder isoleucine in one or two of their repeats. Dependent on the amino acid composition of the repeats, the hypervariable regions were divided into five types. I have not detected any connection between the type of hypervariable region and the similarity of the alleles that has been found by doing two-way hierarchical clustering and the multidimensional scaling. So it can be concluded, that for specificity types of hypervariable regions do not have to be combined with particular patterns of amino acid substitutions in the rest of the gene to give rise to allelic specificity.

Literatutverzeichnis

Adams, J., Rothman, E.D., Kerr, W.E., Paulino, Z.L. (1977). Estimation of the number of sex alleles and queen matings from diploid male frequencies in a population of *Apis mellifera*. (Genetics 86:583–596).

Alberts, B., Bray, C., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (1998). Essential Cell Biology: an introduction to the molecular biology of the cell. Garland Publishing, Inc.

Altschuh, D., Vernet T., Berti P., Moras D. and Nagai K. (1988). Coordinated amino acid changes in homologous protein families. (Protein Eng. 2(3), 193–199).

An J, Totrov M, Abagyan R. Pocketome via comprehensive identification and classification of ligand binding envelopes. Mol Cell Proteomics 2005;4:752–761.

Atchley, W., Wollenberg K.R., Fitch W.M., Terhalle W. and Dress A.W. (2000). Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. (Mol. Biol. Evol. 17(1), 164–178).

Baldwin, E., Hajiseyedjavadi O., Baase W.A. and Matthews B.W. (1993). The role of backbone flexibility in the accommodation of variants that repack the core of T4 lysozyme. (Science 262(5140), 1715–1718).

Bartlett GJ, Porter CT, Borkakoti N, Thornton JM (2002) Analysis of catalytic residues in active sites. J Mol Biol 324: 105–121.

Bashford D, Chothia C, Lesk AM (1987) Determinants of a protein fold. Unique features of the globin amino acid sequences. J Mol Biol 196: 199–216.

Baussand, J. and Carbone, A. (2009). A Combinatorial Approach to Detect Coevolved Amino Acid Networks in Protein Families of Variable Divergence. PLoS Comput Biol 5(9): e1000488. doi:10.1371/journal.pcbi.1000488.

Bell, G. (1982). The Masterpiece of Nature: The Evolution and Genetics of Sexuality. (Croom Helm, London; University of California Press, Berkeley)

Brown, A.J., Casselton, L.A. (2001). Mating in mushrooms: increasing the chances but prolonging the affair. (Trends Genet 17: 393–400).

Beye, M., Hasselmann, M., Fondrk, M.K., Page, R.E. Jr., Omholt, S.W. (2003). The Gene *csd* Is The Primary Signal for Sexual Development in the Honeybee and Encodes an SR-Type Protein. (Cell, Vol. 114, 419-429).

Beye M. (2004). The dice of fate: the *csd* gene and how allelic composition regulates sexual development in the honey bee, *Apis mellifera*. (BioEssays, 26: 1131-1139).

Blencowe, B. J., Bowman, J. A., McCracken, S. and Rosonina, E. (1999). SR-related proteins

and the processing of messenger RNA precursors. (Biochem. Cell Biol. 77, 277–291).

Bordo, D., Argos, P. (1990). Evolution of protein cores. Constraints in point mutations as observed in globin tertiary structure. (J. Mol. Biol. 211(4), 975–988).

Bull, J.J. (1983). Evolution of Sex Determining Mechanisms. (Menlo Park, CA.: Benjamin / Cummings Publishing Company).

Campo, D., Dimitrova, Z., and Khudyakov, Y. (2008). Physicochemical Correlation between Amino Acid Sites in Short Sequences under Selective Pressure. (I. Măndoiu, R. Sunderraman, and A. Zelikovsky (Eds.): ISBRA 2008, LNBI 4983, pp. 146–158, Springer-Verlag Heidelberg).

Cao, W., Jamison, S.F. and Garcia-Blanco, M.A. (1997). Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing *in vitro*. (RNA, 3, 1456–1467.)

Charlesworth, D. (2004). Sex determination: balancing selection in the honey bee. (Curr. Biol. 14: R568-R569).

Chelvanayagam, G., Eggenschwiler A., Knecht L., Gonnet G.H. and Benner S.A. (1997). An analysis of simultaneous variation in protein structures. (Protein Eng. 10(4), 307–316).

Chothia, C., Lesk, A. (1982). Evolution of proteins formed by beta-sheets. I. Plastocyanin and azurin. J. (Mol. Biol. 160(2), 309–323).

Clarke B. (1970). Selective constraints on amino-acid substitutions during the proteins. (Nature, 228(5267):159-160).

Clarke, N. (1995): Covariation of residues in the homeodomain sequence family. (Protein Sci. 4(11), 2269–2278).

Clarke, K., Warwick, R. (2001). Change in marine communities: an approach to statistical analysis and interpretation. (2nd edition. PRIMER-E: Plymouth.)

Cline, T.W. and Meyer, B.J. (1996). Vive la Difference: Males vs females in flies vs worms. (Annu Rev Genet 30:637–702).

Codoner, F., O’Dea, S., Fares, M. (2008). Reducing the false positive rate in the non-parametric analysis of molecular coevolution. (BMC Evolutionary Biology, 8: 106).

Chookajorn, T., Kachroo, A., Daniel R. Ripoll, D.R., Andrew G. Clark, A.G. and Nasrallah, J.B. (2004). Specificity determinants and diversification of the Brassica self-incompatibility pollen ligand. (PNAS, Vol. 101, No. 4, 911–917)

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C. and Duncan, P.I. (1996). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intracellular distribution. (EMBO J., 15, 265–275.)

Cook, J.M. (1993). Sex determination in the hymenoptera: a review of models and evidence. (Heredity 71, 421-435).

Crozier, R.H. (1971). Heterozygosity and sex determination in haplodiploidie. (Am. Nat. 105, 399-412).

Dauwalder, B., and Mattox, W. (1998). Analysis of the functional specificity of RS domains in vivo. (EMBO J 17:6049–6060).

Dayhoff (1978). A model of evolutionary changes in proteins. (*Atlas of protein sequence and structure Volume 5. Issue Suppl. 3. Washington, D.C., National Biomedical Research Foundation; 345-352.*)

Dutheil, J., and Galtier, N. (2007). Detecting groups of coevolving positions in a molecule: a clustering approach. (BMC Evol Biol 2007, 7:242).

Epstein C.J. (1967): Non-randomness of amino-acid changes in the evolution of homologous proteins. (*Nature*, 215(99):355-359).

Fukami-Kobayashi, K., Schreiber, D., Benner, S.. Detecting compensatory covariation signals in protein evolution using reconstructed ancestral sequences. (2002). J. Mol. Biol. 319, 729–743

Galitsky, B. (2003). Revealing the set of mutually correlated positions for the protein families of immunoglobulin fold. (In Silico Biol, 3(3): 241-264.

Gempe, T. (2006). Vollständige Charakterisierung des geschlechtsbestimmenden Lokus der Honigbiene *Apis mellifera*. Inauguraldissertation zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf.

Gempe, T., Hasselmann, M., Schiott, M. Hause, G., Otte, M., Beye, M. (2009). Sex Determination in Honeybees: Two Separate Mechanisms Induce and Maintain the Female Pathway. (PLoS Biol 7(10): e1000222).

Gloor, G., Martin, L.C., Wahl, L.M., Dunn, S.D. (2005): Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions. (Biochemistry 44(19), 156–165).

Goldman, N., Thorne, J., Jones, D. (1998). Assessing the impact of secondary structure and solvent protein evolution. (Genetics, 149(1):445-458).

Goodfellow, P.N and Lovell-Badge, R. (1993). Sry and sex determination in mammals. (Annu. Rev. Genet. 27: 71-92).

Govindarajan, S., Ness, J.E., Kim, S., Mundorff, E.C., Minshull, J. and Gustafsson, C. (2003): Systematic variation of Amino acid substitutions for stringent assessment of pairwise covariation. (J. Mol. Biol. 328, 1061–1069).

Grantham, R. (1974): Amino acid difference formula to help explain protein evolution. (*Science*, 185:862-864).

Graveley, B.R., Hertel, K.J. and Maniatis, T. (1998). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. (*The EMBO Journal* Vol.17 No.22 pp.6747–6756)

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. (*RNA* 6:1197–1211).

Grishin NV, Phillips MA (1994) The subunit interfaces of oligomeric enzymes are conserved to a similar extent to the overall protein sequences. *Protein Sci* 3: 2455–2458.

Grishin, N.V. (2001) Fold change in evolution of protein structures. *J. Struct. Biol.*, 134, 167–185.

Gromiha, M., Oobatake, M., Kono, H., Uedaira, H., Sarai, A. (1999) Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations, *Protein Eng.* 12 549–555.

Gromiha, M., Oobatake, M., Kono, H., Uedaira, H., Sarai, A. (1999). Relationship between amino acid properties and protein stability: buried mutations. *J. Protein Chem.* 18 565–578.

Gromiha, M., Oobatake, M., Kono, H., Uedaira, H., Sarai, A. (2000) Importance of surrounding residues for protein stability of partially buried mutations, *J. Biomol. Struct. Dyn.* 18 281–295.

Hasselmann M, and Beye M. (2004). Signatures of selection among sex-determining alleles of the honey bee. (*Proc Natl Acad Sci USA.* 101:4888–4893).

Hasselmann M, and Beye M. (2006). Pronounced differences of recombination activity at the sex determination locus (SDL) of the honey bee, a locus under strong balancing selection. (*Genetics.* 174:1469–1480).

Hasselmann, M., Fondrk, M.K., Page, R.E., and Beye, M. (2001). Fine scale mapping in the sex locus region of the honey bee (*Apis mellifera*). (*Insect Mol. Biol.* 10, p.605-608)

Hasselmann, M., Vekemans, X., Pflugfelder, J., Koeniger, N., Koeniger, G., et al. (2008 (1)) Evidence for convergent nucleotide evolution and high allelic turnover rates at the complementary sex determiner gene of Western and Asian honeybees. (*Mol Biol Evol* 25: 696–708).

Hasselmann, M., Gempe, T., Schiott, M., Nunes-Silva, C.G., Otte, M., Beye, M. (2008 (2)) Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. (*Nature* 454: 519–522).

Henikoff S., Henikoff J.G. (1992): Amino acid substitution matrices from protein blocks. (*Proc Natl Acad Sci USA* , 89(22):10915-10919).

Hudson, R.R., and Kaplan, N.L. (1988). The coalescent process in models with selection and recombination. (*Genetics*. 120:831–840).

Kahraman A, Morris RJ, Laskowski RA, Thornton JM. (2007). Shape variation in protein binding pockets and their ligands. *J Mol Biol*; 368:283–301.

Kahraman, A., Morris, R.J., Laskowski, A., Faviaa, A.D. and Thornton, J. (2009). On the diversity of physicochemical environments experienced by identical ligands in binding pockets of unrelated proteins. Wiley InterScience, DOI: 10.1002/prot.22633

Kemper, J., Reichmann, M., Romeis, T., Bolker, M., and Kahmann, R. (1995). Multiallelic recognition: nonself-dependent dimerization of the bE and bW homeodomain proteins in *Ustilago maydis*. (*Cell* 81:73–83).

Kay, B.K., Williamson, M.P., and Sudol, M. (2000). The importance of being Proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. (*FASEB J*. 14, 231-241).

Kimura, M. and Crow, J.F. (1964). The number of alleles that can be maintained in a finite population. (*Genetics* 49,: 725-738).

Laidlaw, H.H., Gomes, F.P., and Kerr, W.E. (1956). Estimations of the number of lethal alleles in a panmictic population of *Apis mellifera*. (*Genetics* 41, 179-188).

Lantto, J. and Ohlin, M. (2002) Functional consequences of insertions and deletions in the complementarity-determining regions of human antibodies. *J. Biol. Chem.*, 277, 45108–45114.

Lesk, A., C., C. (1982). Evolution of proteins formed by beta-sheets. II. The core of the immunoglobulin domains. (*J. Mol. Biol.* 160(2), 325–342).

Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci* 1998;7:1884–1897.

Lim, W., Sauer, R. (1989). Alternative packing arrangements in the hydrophobic core of lambda repressor. (*Nature* 339(6219), 31–36).

Lim, W., Farruggio, D., Sauer, R. (1992): Structural and energetic consequences of disruptive mutations in a protein core. (*Biochemistry* 31(17), 4324–4333).

Mac Kean, D.J., Bell, M., and Potter, M. (1978). Mechanisms of antibody diversity: Multiple genes encode structurally related mouse K variable regions. (*Proc. Natl. Acad. Sci. USA*, Vol. 75, No. 8, pp. 3913-3917, Immunology)

- Mackensen, O. (1955). Further studies on a lethal series in the honeybee. (*J. Hered.* 46: 72–74).
- Martin, L., Gloor, G.B., Dunn, S.D., Wahl, L.M.. (2005). Using information theory to search for co-evolving residues in proteins. (*Bioinformatics* 21 (22), 4116–4124).
- Matthews, B.W. (1987) Genetic and structural analysis of the protein stability problem. *Biochemistry*, 26, 6885–6888.
- Nasrallah, J.B. (2002). Recognition and rejection of self in plant reproduction. (*Science* 296: 305–308).
- McClure, B. (2004). S-RNase and SLF determine S-haplotype-specific pollen recognition and rejection. (*Plant Cell* 16: 2840–2847).
- Mateu, M., Fersht, A. (1999). Mutually compensatory mutations during evolution of the tetramerization domain of tumor suppressor p53 lead to impaired hetero-oligomerization. (*Proc. Natl. Acad. Sci. USA* 96, 3595–3599).
- Miyata, T., Miyazawa, S., Yasunaga, T. (1979). Two Types of Amino Acid Substitutions in Protein Evolution. (*J. Mol. Evol.* 12, 219-236)
- Neher, E. (1994). How frequent are correlated changes in families of protein sequences? (*Proceedings of the National Academy of Sciences of the United States of America*, 91:98-102.)
- Oishi, K., Sawa, M., Hatakeyama, M., and Kageyama, Y. (1993). Genetics and biology of the sawfly, *Atalia rosae* (Hymenoptera). (*Genetics* 88, 119-127).
- Oosawa, K., Simon, M. (1986). Analysis of mutations in the transmembrane region of the aspartate chemoreceptor in *Escherichia coli*. (*Proc. Natl. Acad. Sci. USA* 83(18), 6930–6934).
- Orr HA: The population genetics of adaptation: the adaptation of DNA sequences. *Evolution* 2002, 56:1317-1330.
- Osuna, J., Jorge Ya'nez, J., Sobero'n, X. and Gayta'n, P.(2004). Protein evolution by codon-based random deletions. *Nucleic Acids Research*, 2004, Vol. 32, No. 17, e136.
- Ouzounis C, Perez-Irratxeta C, Sander C, Valencia A (1998) Are binding residues conserved? *Pac Symp Biocomput* 3: 401–412.
- Padlan, E. (1977). Structural basis for the specificity of antibody-antigen reactions and structural mechanisms for the diversification of antigen-binding specificities. (*Quant. Rev. Biophys.* 10, 35-65).

Padlan, E., Davis, D. (1975). Variability of three-dimensional structure in immunoglobulins. (Proc. Nat. Acad. Sci. USA 72, 819-823).

Pascarella, S. and Argos, P. (1992) Analysis of insertions/deletions in protein structures. J. Mol. Biol., 224, 461–471.

Pear, J.R., Kawagoe, Y., Schreckengost, W.E., Delmer, D.P. and Stalker, D.M. (1996). Higher plants contain homologs of the bacterial *ce1A* genes encoding the catalytic subunit of cellulose synthase (Proc. Natl. Acad. Sci. USA, Vol. 93, pp. 12637-12642, Plant Biology)

Pincus, J.H., Haber, E., Katz, M. and Pappenheimer, A.M. (1968). Antibodies to pneumococcal polysaccharides: Relation between binding and electrophoretic heterogeneity. (Science (Washington) 162: 667.)

Poljak, R., Amzel, L., Phizackerley, R. (1976). Studies on the three-dimensional structure of immunoglobulins. (Prog. Biophys. Molec. Biol. 31, 67-93).

Pollock, D., Taylor, W. (1997). Effectiveness of correlation analysis in identifying protein residues. (Protein Eng. 10(6), 647–657).

Poon, A., Chao, L. (2005). The rate of compensatory mutation in the DNA bacteriophage phiX174. (Genetics 170(3), 989–999).

Rudikoff, S., Rao, D. N., Glaudemans, C. P. J. and Potter, M. (1980). (Proc. Natl Acad. Sci. USA 77, 4270-4274).

Rudikoff, S., Giusti, A.M., Cook, W.D., and Scharff, M.D. (1982). Single amino acid substitution altering antigen-binding specificity. (Proc. Nat. Acad. Sci. USA, Vol. 79, pp. 1979-1983, Immunology)

Schmitt A, Schuchhardt J, Brockmann GA (2009) The Action of Key Factors in Protein Evolution at High Temporal Resolution. PLoS ONE 4(3): e4821.
doi:10.1371/journal.pone.0004821

Shen, B., Bai, J., Vihinen, M. (2008). Physicochemical feature-based classification of amino acid mutations. (Protein Engineering, Design & Selection vol. 21 no. 1 pp. 37-44)

Shortle, D. and Sondek, J. (1995) The emerging role of insertions and deletions in protein engineering. Curr. Opin. Biotechnol., 6, 387–393.

Sondek, J. and Shortle, D. (1992) Structural and energetic differences between insertions and substitutions in staphylococcal nuclease. Proteins, 13, 132–140.

Stone, E.A. and Sidow, A. (2005). Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. 15:978–986
©2005 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/05

Stryer, L. (1995). Biochemistry. W.H. Freeman and Company, New York.

Sudol M, Sliwa K, Russo T. 2001. Functions of WW domains in the nucleus. FEBS Lett 490:190–195.

Suel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionary conserved networks of residues mediate allosteric communication in proteins. Nature Struct Biol 23: 59–69.

Tacke, R., and Manley, J.L. (1999). Determinants of SR protein specificity. (Curr. Opin. Cell Biol. 11:358–362).

Takahata, N. (1990). A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. (Proc Natl Acad Sci USA. 87:2419–2423).

Trent, C., Crosby, C. and Eavey, J. (2006). Additional evidence for the genomic imprinting model of sex determination in the haplodiploid wasp *Nasonia vitripennis*: isolation of biparental diploid males after X-ray mutagenesis. (Heredity 96, 368-376).

Tsai C-J, Norel R, Wolfson HJ, Maizel JV, Nussinov R. Protein- Ligand Interactions: Energetic Contributions and Shape Complementarity. In: Encyclopedia of Life Sciences. John Wiley & Sons, Ltd: Chichester <http://www.els.net/> [doi: 10.1038/npg.els.0001343].

Wang H., Lin, W., Dyck, J. A., Yeakley, J.M., Songyang, Z., Cantley, L.C. and Fu, X. (1998). SRPK2: A Differentially Expressed SR Protein-specific Kinase Involved in Mediating the Interaction and Localization of Pre-mRNA Splicing Factors in Mammalian Cells. (The Journal of Cell Biology, Volume 140, Number 4, 737–750)

Wang, J., Xiao, S.H., and Manley, J.L. (1998). Genetic analysis of the SR protein ASF/SF2: interchangeability of RS domains and negative control of splicing. (Genes Dev 12:2222–2233).

Whiting, P.W. (1933). Selective fertilization and sex-determination in Hymenoptera. (Science 78: 537–538).

Whiting, P.W. (1943). Multiple alleles in complementary sex determination of *Habrobracon*. (Genetics 28: 365–382).

Wong, W., Sainudiin, R., Nielsen, R. (2006). Identification of physicochemical selective pressure on Protein encoding nucleotide sequences. (BMC Bioinformatics, 7: 148).

Xiao, S.-H. and Manley, J.L. (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein–protein interactions and is necessary for splicing. (Genes Dev., 11, 334–344.)

Yeakley, J.M., Tronchere, H., Olesen, J., Dyck, J.A., Wang, H.Y., et al. (1999). Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors. (J Cell Biol 145:447– 455).

Yeang, C. and Haussler, D. (2007): Detecting coevolution in and among protein domains. (PLoS Comput Biol. 3(11), e211).

Yokoyama, S. and Nei, M. (1979). Population dynamics of sexdetermining alleles in honey bees and self-incompatibility in plants. (Genetics. 91:609–626).

Zheng W, Brooks BR, Doniach S, Thirumalai D (2005) Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. Structure 13: 565–577.

Primersequenzen und Annealingtemperaturen

Tabelle A.I: Primer für Klonierungen und Genotypisierungen (Restriktionsanalysen) von Exon 6-9 Sequenzen von *csd* der F1- und F2-Drohnen

forward-Primer: konservierte Sequenzen

fw-Primer	Nukleotisequenz	Annealing-temperatur	Exon
5_Kag1_210_fw	TCC TGT TCC TGT TCC TGT TTA TTA TG	50°C	7
Geno_rnai_1b_fw	CAGACAATATGAAAAATTACACAATGA	49°C	6
csdE6-for-1	AATGAAAAAGAAAACTTTTAGAAGAAAGAAC	49°C	6
GenoInt_fw1	(CT)AGAC(AG)ATATGAAAAATTA(CT)A(CT)AA(CT)	51°C	6
GenoInt_fw2	A(CT)AA(CT)GAA(AG)AAGAAAAA(CT)TTTT	51°C	6
SIQ30_82_fw	AGAACGAGAGCAAAAAGTATG	50°C	6
consBGRWY_123_fw 4	CATATTT(GA)TTAAA(AC)ATTGAATTTTAATTTA	50°C	Intron 6

reversed-Primer:

rev-Primer	Nukleotisequenz	Annealing-temperatur	Exon
IGconsrev-1	TAAAATTTTTATAGTTTTTCATTGATGCGTAG	49°C	9
IGconsrev-2	TGACAGATTTTACATTTCTATTTTTTG	49°C	9
Cons_CSD_rev	(AG)TC ATC TCA T(AT)T TTC ATT ATT CAA T	49°C	9
SIQ30_495_rv	TTGGTGGAAAATTCTAATAATG	50°C	7
Csd_cons_rev07	CAATATGT(CT)GCATCAACTTTT	51°C	9
spezlc_rev	TCCAAGGTCCCATTTGGTCTT	49°C	8
consBGRWY_123_rev	TCA TTG ATG CGT AGG TCC AAA TCC T	50°C	9

Tabelle A.II. forward-Primer für Klonierungen der 5'-UTR bis Exon 6/7 Sequenzen von *csd*

fw-Primer	Nukleotisequenz	Annealingtemperatur	Exon	Spezifität
cons_csd_for	GGTGATTATACATTTGCAGGT	48°C	1	konserviert
cons_csd_for5	TGGGTGATTATACGTTAAATTTCA	48°C	1	konserviert
fw_CSD	CTTGTTCCGGTATT(TA)TCATAAA	48°C	1	konserviert

Tabelle A. III: reversed-Primer reversed-Primer mit deren Hilfe in Kombination mit konservierten fw-Primern (Tab. AII) gezielt jeweils nur eines der im cDNA-Pool enthaltenen *csd*-Allele amplifiziert werden konnte.

rev-Primer	Allel	Nukleotisequenz	Annealingtemperatur	Exon
B1a_rev	B1	TTATTATAATTGCTATATTTGTAT	48°C	7
B2a_rev	B2	TAGAAATGTTAGAAATGTAATTG	48°	7
B3c_rev	B3	TAATTATTATTATTATGTATTGTCT	48°	7
G1a_rev	G3	GATGGACCAATATGCCTAAATCTTGG	48°	8
G3a_rev	G3	CAAAGGTCCCATTGGTCTTG	48°	8
R1a_rev	R1	AATTAGAAATGTTAGAAATGTAATTGTTC	48°	7
R2b_rev	R2	GGTGGAAAATTTCCATAATAC	48°	7/8
R2_rev2b	R2	GATATTCTCTTTCATTTTTATATGAC	48°	6/7
R3b_rev	R3	TTCTCGATATTTTCGATAAGAA	48°	7
consW_1234_rev3	W1	AAATCTTGGTATTTGTTCTTG	48°	8
consW_1234_rev2	W2	TCTCGTTCT(GC)TTCTATCTC	48°	7
W13_rev2	W3	ATTGCTATAATTGCTATAATTG	48°	7
Y1_3a_rev	Y1, Y3	TCTCGATA(TC)TTTCGATAAGAAT	48°	7
Y2a_rev	Y2	TCTCCATATTTTCGATATTCTC	48°	7

Tabelle A.IV. RACE-Primer

rev-Primer	Allel	Nukleotisequenz	Annealing-temperatur	Exon
G1_2_outer_rev	G1, G2	ATAAATAGGAACAGGAACAGGAAC	60°C	7
G1_inner_rev	G1	TTCTTTAGATCTCTCTCGTTCCGT	60°C	7
G2_inner_rev	G2	GTTCTTTAGATCTCTCTCGTTCTC	60°C	7

Tabelle A.V: Primer für radioaktive PCR (Fragmentgröße ca. 615 bp)

Primer	Allel	Nukleotisequenz	Annealing-temperatur	Exon/Intron
cons_B123_fw	alle	CACAATGAAAAAGAAAACTTTT	50°C	Exon 6
cons_G123_rev	alle	ACTATGTGCATCAATATAAATTC	50°C	Intron 8

Abbildung A.I: Alignment der Aminosäuresequenzen der 14 *csd* Allele: Die Aminosäuren wurden mit dem Einbuchstabencode abgekürzt, fehlende Aminosäuren (Deletionen) sind mit - gekennzeichnet.

		10	20	30	40	50
G1		MKRNISSYSH	HDEKFKQLRN	EDSEIDLRSR	TKEERLQHRR	EAWLIQQERE
Y2		MKRNISSYSH	HDEKFKQLRN	EDNKIDLRSR	TKEERLQHRR	EVWLIQQERE
B1		MKRNISSYSH	HDEKFKQLRN	EDSEIDLRSR	TKEERLQHRR	EAWLIQQERE
G2		MKRNISSYSH	HDEKFKQLRN	EDNKIDLRSR	TKEERLQHRR	EVWLIQQERE
W1		MKRNISSYSH	HDEKFKQLRN	EDSEIDLRSR	TKEERLQHRR	EAWLIQQERE
G3		MKRNISSYSH	HDEKFKQLRN	EDSEIDLRSR	TKEERLQHRR	EAWLIQQERE
R1		MKRNISSYSH	HDEKFKQLRN	EDNKIDLRSR	TKEERLQHRR	EVWLIQQERE
B2		MKRNISSYSH	HDEKFKQLRN	EDNKIDLRSR	TKEERLQHRR	EVWLIQQERE
W2		MKRNISSYSH	HDEKFKQLRN	EDNKIDLRSR	TKEERLQHRR	EVWLIQQERE
R3		MKRNISSYSH	HDEKFKQLRN	EDNKIDLRSR	TKEERLQHRR	EAWLIQQERE
Y1		MKRNISSYSH	HDEKFKQLRN	EDSEIDLRSR	TKEERLQHRR	EAWLIQQERE
R2		MKRNISSYSH	HDEKFKQLRN	EDSEIDLRSR	TKEERLQHRR	EAWLIQQERE
Y3		MKRNISSYSH	HDEKFKQLRN	EDNKIDLRSR	TKEERLQHRR	EVWLIQQERE
B3		MKRNISSYSH	HDEKFKQLRN	EDNKIDLRSR	TKEERLQHRR	EVWLIQQERE

		60	70	80	90	100
G1		REHQRLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNASNTSKTV
Y2		REHERLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNTSNTSKTV
B1		REHQRLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNTSNTSKTV
G2		REHERLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNTSNTSKTV
W1		REHQRLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNTSNTSKTV
G3		REHQRLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNASNTSKTV
R1		REHERLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNTSNTSKTV
B2		REHERLIKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNTSNTSKTV
W2		REHERLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNTSNTSKTV
R3		REHERLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNTSNTSKTV
Y1		REHQRLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNTSNTSKTV
R2		REHQRLMKKM	ILEYELRRIR	EIEKLGSESR	KSTSPDSRDR	SNASNTSKTV
Y3		REHERLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNTSNTSKTV
B3		REHERLMKKM	ILEYELRRIR	EIEKLGSESR	KSRSPDSRDR	SNTSNTSKTV

		110	120	130	140	150
G1		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
Y2		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
B1		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
G2		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
W1		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
G3		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
R1		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKS-TTTTVEV
B2		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
W2		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
R3		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
Y1		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
R2		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
Y3		ILSNKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV
B3		ILSDKLESSD	DISLFRGPEG	IQINATELQK	IKLEIHRDLP	GKSTTTTVEV

	160	170	180	190	200
G1	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIE	EHDTVLLVNI
Y2	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIE	EHDTVLLVNI
B1	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIE	EHDTVLLVNI
G2	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIK	EHDTVLLVNI
W1	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIK	EHDTVLLVNI
G3	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIE	EHDTVLLVNI
R1	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIK	EHDTVLLVNI
B2	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIE	EHDTVLLVNI
W2	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIK	EHDTVLLVNI
R3	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIK	EHDTVLLVNI
Y1	KRDIINPEDV	ILIRRTGEGS	KPLFEREEIK	NVLTKINKIE	EHDTVLLVNI
R2	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIK	EHDTVLLVNI
Y3	KRDIINPEDV	IVIRRTGEGS	KPLFEREEIK	NVLTKINKIE	EHDTVLLVNI
B3	KRDIINPEDV	ILIRRTGEGS	KPIFEREEIK	NVLTKINKIE	EHDTVLLVNI

	210	220	230	240	250
G1	KKSGNESKKY	ATSSNSLRSR	THGFQHTSSR	YSRERSCSR	RNREYRKKDR
Y2	KKSGNESKKY	ATSSNSLRSR	THGFQHTSSR	YSRERSCSR	RNREYRKKDR
B1	KKSGNESKKY	ATSSNSLRSR	THGFQHTSSR	YSRERSCSR	RNREYRKKDR
G2	EKSGNESKKY	ATSSNSLRNR	THGFQHTSSR	YSRERSCSR	RNREYKEKDR
W1	EKSGNESKKY	ATSSNSLRNR	THGFQHTSSR	YSRERSYSR	RNREYRKKDR
G3	KKSGNESKKY	ATSSNSLRSR	THGFQHTSSH	YSRERSCSR	RNREYRKKDR
R1	EKSGNESKKY	ATSSNSLRNR	THGFQHTSSR	YSRERSCSR	RNREYKEKDR
B2	KKSGNESKKY	ATSSNSLRSR	THGFQHTSSR	YSRERSCSR	RNREYKEKDR
W2	EKSGNESKKY	-TSSNSLRNR	IHGFQHTSSG	YSRERSCSR	RNREYRKKDR
R3	EKSGNESKKY	ATSSNSLRNR	THGFQHTSSR	YSRERSCSR	RNREYKEKDR
Y1	KKSGNESKKY	ATSSNSLRNR	THGFQHTSSR	YSRERSCSR	RNREYKEKDR
R2	EKSGNESKKY	ATSSNSLRNR	THGFQHTSSR	YSRERSCSR	RNREYRKKDR
Y3	EKSGNESKKY	ATSSNSLRNR	THGFQHTSSR	YSRERSCSR	RNREYKEKDR
B3	KKSGNESKKY	ATSSNSLRSR	THGFQHTSSR	YSRERSCSR	RNREYKEKDR

	260	270	280	290	300
G1	RYEKLHNEKE	KLLEERTSRE	RYRSREREQ	KSYKNEREYR	KYGETSKERS
Y2	RYEKLHNEKE	KLLEERTSRE	RYRSREREQ	KSYKNEREYR	KYGETSKERS
B1	QYEKLHNEKE	KLLEERTSRK	RYRSREREQ	KSYKNEREYR	KYGETSKERS
G2	RYEKLHNEKE	KLLEERTSRK	RYRSREREQ	NSYKNEREYR	KYRETSKERS
W1	RYEKLHNEKE	KLLEERTSRK	RYRSREREQ	KSYKNENSYR	KYRETSKERS
G3	RYEKLHNEKE	KLLEERTSCK	RYRSREREQ	KLYKNEREYR	KYGETSKERS
R1	RYEKLHNEKE	KLLEERTSRK	RYRSREREQ	NSYKNEKEYR	KYRERSKERS
B2	RYEKLHNEKE	KLLEGRTSRK	RYRSREREQ	KSYKNENSYR	KYRETSRERS
W2	QYEKLCNEEE	KLLEERTSRK	RYRSREREQ	NSYKNEREYR	KYRETSKERS
R3	RYEKLHNEKE	KLLEERTSRK	RYRSREREQ	KSYKNENSYR	KYRERSRERS
Y1	RYEKLHNEKE	KLLEERTSRK	RYRSREREQ	KSYKNENSYR	KYRETSKERS
R2	RYEKLHNEKE	KLLEERTSRE	RYRSREREQ	KSYKNEREYR	KYRETSKERS
Y3	RYEKLHNEKE	KLLEERTSRK	RYFRSREREQ	KSYKNENSYR	KYRETSKERS
B3	RYEKLHNEKE	KLLEERTSRK	RYRSREREQ	NSYKNEREYR	KYRERSKERS

	310	320	330	340	350
G1	RDRTERERSK	EPKIISSLSN	NYKYSNYNNY	NNNYNNYNNY	NNNYNNNY--
Y2	RDRTERERSK	EPKIISSLSN	NYKYSNYNNY	NNNYNNYNNY	NNNYNNYNNN
B1	RDRTERERSK	EPKIISSLSN	NYKYSNYNNY	NNYNNNS---	-----
G2	RDRTERERSK	EPKIISSLSN	NYNYSNYNNY	NNNYNNY--	-----
W1	RDRTERERSK	ESKIISSLSN	NYNYNNCNYK	H-----	-----
G3	RNRTEREKSK	EPKIISSLSN	NYNYNNYNNN	Y-----	-----
R1	RDKRERERSK	ERKIISSLSN	NYISNISNYN	NNNNY--	-----
B2	RDRKER-RSK	ERKIISSLSN	NYISNISNYN	NNNS-----	-----
W2	RDRTERERSR	EPKIISSLSN	NTIHNNNYKY	NYNNNNY--	-----
R3	RDRTERERSK	EPKIISSLSN	KTIHNNNYK	YNYKNYNS-	-----
Y1	RDRTERERSR	EPKIISSLSN	KTIHNNNYK	YNYNNC---	-----
R2	RDRTERERSR	EPKIISSLSN	KTIHNNNYK	YNYNNNNNY	KNYNNY----
Y3	RDRTERERSK	EPKIISSLSN	KTIHNNNYN	NNYNNNYKNY	NNS-----
B3	RDRTERERSR	EPKIISSLSN	KTIHNNNYN	NNNYNNYNN	YNNNNYNNY-

	360	370	380	390	400
G1	-KKLYN--I	NYIEQIPVPV	PVPIYCGNFP	PRSMEPWISM	QEQIPRFRHI
Y2	YKKLYN--I	NYIEQIPVPV	PVPIYCGNFP	PRSMEPWISM	QEQIPRFRHI
B1	-KKLYKNI	N-IEQIPVPV	PIPVYCGNFP	PRPMGPWISM	QEQIPRFRHI
G2	-KKLYN--I	NYIEQIPVPV	PVPIYCGNFP	PRPMGPWISM	QEQIPRFRYI
W1	-NKLYN-II	N-IEQIPVPV	P--IYCGNFP	PRPMGPWFMS	QEQIPRFRHI
G3	-KPLYN--I	NYIEQIPVPV	PFVYVYGNFP	PRPMGPFVSI	QEQIPRFRYI
R1	-KKLYN--I	NYIEQIPVPV	PVPIYCGNFP	PRPMGPWISI	QEQIPRFRYI
B2	-KKLYN--I	NYIEQIPVPI	PVPVYCGNFP	PRPMGPWISM	QEQIPRFRYI
W2	-KKLYN--I	NYIEQIPIV	P--VYVGNFL	PRPMGPWISI	QEQIPRFRHI
R3	-KKLYN-VI	N-IEQIPVPV	PVPIYCGNFP	PRPMGPWISI	QEQIPRFRYI
Y1	-KKLYN--I	NYIEQIPIV	P--VYVGNFP	PRPMGPWISM	QEQIPRFRYI
R2	-KKLYN--I	NYIEQIPVPV	P--VYVGNFP	PRPMGPWISI	QEQIPRFRHI
Y3	-KKLYN--I	NYIEQIPIV	PLPIYCGNFP	PRPMGPWFMS	QEEIPRFRYI
B3	-KKLYN-II	N-IEQIPVPV	PVPIYCGNFP	PRPMGPWISI	QEQIPRFRYI

	410	420	430
G1	GPSTPFPRFI	PPNAYR-FRP	PLNPRFGPTH Q
Y2	GPSTPFPRFI	PPNAYR-FRP	PLNPRFGPTH Q
B1	GPSTPFPRFI	SPNAYRYIFP	PLNPRFGPTH -
G2	GPSTPFPRFI	PPNAYR-FRP	PLNPRFGLAH Q
W1	GPSTPFPRFI	SPNAYR-FRP	PLNPRFGPTH Q
G3	GPSTPFPRFI	PPNAYR-FRP	SLNPRFGSTH Q
R1	GPSTPFPRFI	PPNAYR-FRP	PQNPRFGPTH Q
B2	GPSTPFPRFI	PPNAYR-FRP	PLNPRFEPTH Q
W2	GPSTPFPRFI	PPNAYR-FRP	PLNPRFEPTH -
R3	GPSTPFPRFI	PPNAYR-FRP	PLNPRFGPTH Q
Y1	GPSTPFPRFI	PPNAYR-FHP	PLNPRFGPTH Q
R2	GPSTPFPRFI	PPNAYR-FRP	PLNPRFGPTH Q
Y3	GPSTPFPRFI	SPNAYR-FRP	PLNPRFGPTH Q
B3	GPSTPFPRFI	PPNAYR-FRP	PLNPRFGPTH Q

Tabelle A.VI: Matrix1: Angegeben sind die Positionswerte der 82 Positionen bei den 14 Allelen

	7	23	24	42	54	57	83	93	104	144	162	173	190	201	205	211	219	221	230	237	246	247	251	256	259	265	267	269	270	
B1	0,7 3	0,7 3	0,7 3	0,6 4	0,7 3	0,0 9	0,0 9	0,1 8	0,0 9	0,0 9	0,0 9	0,0 9	0,5 5	0,6 4	0,0 9	0,0 9	0,6 4	0,0 9	0,1 8	0,0 9	0,5 5	0,5 5	0,9 1	0,0 9	0,0 9	0,0 9	0,0 9	0,0 9	0,0 9	0,1 8
B2	0,4 5	0,4 5	0,4 5	0,5 5	0,4 5	1,0 0	0,0 9	0,2 7	0,0 9	0,0 0	0,0 9	0,0 9	0,4 5	0,5 5	0,0 9	0,0 9	0,5 5	0,0 9	0,1 8	0,0 9	0,6 4	0,6 4	0,1 8	0,0 9	0,0 9	1,0 0	0,0 9	0,0 9	0,0 9	0,2 7
B3	0,5 0	0,5 0	0,5 0	0,6 7	0,5 0	0,1 7	0,0 0	0,3 3	0,0 0	0,1 7	0,0 0	0,0 0	0,3 3	0,3 3	0,0 0	0,0 0	0,3 3	0,0 0	0,1 7	0,0 0	0,5 0	0,5 0	0,1 7	0,0 0	0,0 0	0,1 7	0,0 0	0,1 7	0,1 7	
G1	0,6 7	0,6 7	0,6 7	0,6 7	0,6 7	0,1 1	0,1 1	0,7 8	0,1 1	0,1 1	0,1 1	0,1 1	0,4 4	0,5 6	0,1 1	0,0 0	0,5 6	0,0 0	0,1 1	0,1 1	0,5 6	0,5 6	0,0 0	0,0 0	0,0 0	0,1 1	0,1 1	0,1 1	0,7 8	
G2	0,5 0	0,5 0	0,5 0	0,5 8	0,5 0	0,0 8	0,0 8	0,2 5	0,0 8	0,0 8	0,0 8	0,1 7	0,5 8	0,5 0	0,0 8	0,0 8	0,4 2	0,0 8	0,1 7	0,0 8	0,5 8	0,5 8	0,1 7	0,0 8	0,0 8	0,0 8	0,0 8	0,0 8	0,0 8	0,2 5
G3	0,5 8	0,5 8	0,5 8	0,5 0	0,5 8	0,0 8	0,0 8	0,8 3	0,0 8	0,0 8	0,0 8	0,1 7	0,4 2	0,5 0	0,0 8	0,0 0	0,5 8	0,0 0	1,0 0	0,0 8	0,5 8	0,5 8	0,0 8	0,0 0	0,0 0	0,0 8	0,0 8	1,0 0	0,2 5	
R1	0,5 5	0,5 5	0,5 5	0,6 4	0,5 5	0,0 0	0,0 9	0,2 7	0,0 9	1,0 0	0,0 9	0,1 8	0,6 4	0,5 5	0,0 9	0,0 0	0,4 5	0,0 0	0,0 9	0,0 9	0,5 5	0,5 5	0,0 9	0,0 0	0,0 0	0,0 0	0,0 0	0,0 9	0,0 9	0,2 7
R2	0,6 0	0,6 0	0,6 0	0,5 0	0,6 0	0,1 0	1,0 0	0,8 0	0,0 0	0,1 0	0,0 0	0,1 0	0,6 0	0,6 0	0,0 0	0,1 0	0,5 0	0,1 0	0,2 0	0,0 0	0,5 0	0,5 0	0,2 0	0,1 0	0,1 0	0,1 0	1,0 0	0,1 0	0,8 0	
R3	0,4 5	0,4 5	0,4 5	0,5 5	0,4 5	0,0 9	0,0 9	0,1 8	0,0 9	0,0 9	0,0 9	0,1 8	0,5 5	0,4 5	0,0 9	0,0 9	0,3 6	0,0 9	0,1 8	0,0 9	0,4 5	0,4 5	0,1 8	0,0 9						
W1	0,6 0	0,6 0	0,6 0	0,5 0	0,6 0	0,1 0	0,0 0	0,2 0	0,0 0	0,1 0	0,0 0	0,1 0	0,6 0	0,6 0	0,0 0	0,1 0	0,5 0	0,1 0	0,2 0	1,0 0	0,5 0	0,5 0	0,2 0	0,1 0	0,1 0	0,1 0	0,1 0	0,0 0	0,1 0	0,2 0
W2	0,3 8	0,3 8	0,3 8	0,5 0	0,3 8	0,1 3	0,1 3	0,1 3	0,1 3	0,0 0	0,1 3	0,2 5	0,6 3	0,5 0	0,1 3	1,0 0	0,3 8	1,0 0	1,0 0	0,0 0	0,6 3	0,6 3	0,8 8	1,0 0	1,0 0	0,1 3	0,1 3	0,0 0	0,2 5	
Y1	0,6 7	0,6 7	0,6 7	0,4 4	0,6 7	0,0 0	0,1 1	0,2 2	0,1 1	0,1 1	0,1 1	0,8 9	0,6 7	0,7 8	0,1 1	0,1 1	0,2 2	0,1 1	0,2 1	0,1 1	0,5 6	0,5 6	0,1 1	0,1 1	0,1 1	0,0 0	0,1 1	0,1 1	0,2 2	
Y2	0,5 5	0,5 5	0,5 5	0,4 5	0,5 5	0,0 9	0,0 9	0,2 7	0,0 9	0,0 9	0,0 9	0,1 8	0,4 5	0,5 5	0,0 9	0,0 9	0,6 4	0,0 9	0,1 8	0,0 9	0,4 5	0,4 5	0,1 8	0,0 9	0,0 9	0,0 9	0,0 9	0,0 9	0,0 9	0,8 2
Y3	0,4 0	0,4 0	0,4 0	0,5 0	0,4 0	0,1 0	0,0 0	0,2 0	1,0 0	0,1 0	1,0 0	0,9 0	0,4 0	0,6 0	1,0 0	0,1 0	0,5 0	0,1 0	0,2 0	0,0 0	0,5 0	0,5 0	0,2 0	0,1 0	0,1 0	0,1 0	0,0 0	0,1 0	0,2 0	

	273	281	282	287	288	293	295	297	302	303	304	305	307	308	310	312	321	322	323	324	352	353	356	358	359	362	368	370	372	
B1	0,0 9	0,3 6	0,0 9	0,4 5	0,3 6	0,8 2	0,2 7	0,1 8	0,0 9	0,0 9	0,3 6	1,0 0	0,0 9	0,0 9	0,2 7	0,2 7	0,3 6	0,4 5	0,9 1	0,6 4	0,0 9	0,0 9	1,0 0	1,0 0	0,8 2	0,7 3	0,1 8	0,0 9	1,0 0	
B2	0,0 9	0,2 7	0,0 9	0,7 3	0,7 3	0,3 6	0,1 8	0,9 1	0,0 9	0,0 0	1,0 0	0,0 9	1,0 0	0,0 9	0,2 7	1,0 0	0,3 6	0,4 5	0,5 5	1,0 0	0,0 9	0,0 9	0,0 9	0,0 9	0,3 6	0,3 6	0,1 8	1,0 0	0,5 5	
B3	0,0 0	0,8 3	0,1 7	0,5 0	0,3 3	0,5 0	0,6 7	0,3 3	0,1 7	0,1 7	0,3 3	0,1 7	0,1 7	0,1 7	1,0 0	0,3 3	0,8 3	0,8 3	0,5 0	0,8 3	0,0 0	0,1 7	0,1 7	0,1 7	0,8 3	0,6 7	0,0 0	0,1 7	0,3 3	
G1	0,1 1	0,3 3	0,1 1	0,4 4	0,3 3	0,7 8	0,2 2	0,1 1	0,1 1	0,1 1	0,4 4	0,0 0	0,1 1	0,1 1	0,2 2	0,3 3	0,3 3	0,3 3	0,8 9	0,5 6	0,1 1	0,1 1	0,0 0	0,0 0	0,2 2	0,2 2	0,1 1	0,1 1	0,4 4	
G2	0,0 8	0,8 3	0,0 8	0,5 0	0,4 2	0,3 3	0,1 7	0,1 7	0,0 8	0,0 8	0,8 3	0,0 8	0,0 8	0,0 8	0,2 5	0,2 5	0,3 3	0,4 2	0,8 3	0,5 8	0,0 8	0,0 8	0,0 8	0,0 8	0,2 5	0,2 5	0,2 5	0,0 8	0,5 8	
G3	0,0 8	0,2 5	1,0 0	0,5 0	0,4 2	0,7 5	0,2 5	0,1 7	1,0 0	0,0 8	0,3 3	0,0 8	0,0 8	1,0 0	0,2 5	0,2 5	0,4 2	0,4 2	0,8 3	0,5 8	0,0 8	1,0 0	0,0 8	0,0 8	0,3 3	0,3 3	0,1 7	0,0 8	1,0 0	
R1	0,0 9	0,8 2	0,0 9	1,0 0	0,3 6	0,3 6	0,8 2	0,0 9	0,0 9	1,0 0	0,8 2	0,0 9	0,0 0	0,0 9	0,2 7	1,0 0	0,4 5	0,4 5	0,5 5	1,0 0	0,0 9	0,0 9	0,0 9	0,0 9	0,3 6	0,3 6	0,1 8	0,0 0	0,5 5	
R2	0,0 0	0,3 0	0,1 0	0,4 0	0,3 0	0,4 0	0,2 0	0,2 0	0,1 0	0,1 0	0,3 0	0,1 0	0,1 0	0,1 0	0,8 0	0,2 0	0,8 0	0,7 0	0,5 0	0,7 0	0,0 0	0,1 0	0,1 0	0,1 0	0,1 0	0,2 0	0,2 0	0,2 0	0,1 0	0,8 0
R3	0,0 9	0,3 6	0,0 9	0,6 4	0,6 4	0,1 8	0,8 2	0,9 1	0,0 9	0,0 9	0,3 6	0,0 9	0,0 9	0,0 9	0,3 6	0,2 7	0,6 4	0,5 5	0,3 6	0,5 5	0,0 9	0,0 9	0,0 9	0,0 9	1,0 0	0,7 3	0,2 7	0,0 9	0,6 4	
W1	0,0 0	0,3 0	0,1 0	0,7 0	0,7 0	0,4 0	0,2 0	0,2 0	0,1 0	0,1 0	0,8 0	0,1 0	0,1 0	0,1 0	0,2 0	1,0 0	0,2 0	0,3 0	0,8 0	0,5 0	1,0 0	0,1 0	0,1 0	0,1 0	0,1 0	0,9 0	0,8 0	0,2 0	0,1 0	0,8 0
W2	0,1 3	0,8 8	0,0 0	0,5 0	0,5 0	0,2 5	0,1 3	0,2 5	0,0 0	0,0 0	0,2 5	0,1 3	0,1 3	0,0 0	0,7 5	0,1 3	0,5 0	0,5 0	0,3 8	0,5 0	0,0 0	0,0 0	0,1 3	0,1 3	0,2 5	0,2 5	0,7 5	0,1 3	0,7 5	
Y1	0,1 1	0,3 3	0,1 1	0,6 7	0,6 7	0,2 2	0,2 2	0,1 1	0,1 1	0,1 1	0,3 3	0,0 0	0,0 0	0,1 1	0,7 8	0,2 2	0,6 7	0,5 6	0,4 4	0,5 6	0,1 1	0,1 1	0,0 0	0,0 0	0,2 2	0,2 2	0,7 8	0,0 0	0,6 7	
Y2	0,0 9	0,2 7	0,0 9	0,4 5	0,3 6	0,7 3	0,0 9	0,0 9	0,0 9	0,0 9	0,3 6	0,0 9	0,0 9	0,0 9	0,2 7	0,2 7	0,2 7	0,3 6	0,8 2	0,5 5	0,0 9	0,0 9	0,0 9	0,0 9	0,1 8	0,1 8	0,2 7	0,0 9	0,6 4	
Y3	1,0 0	0,3 0	0,1 0	0,7 0	0,7 0	0,4 0	0,2 0	0,2 0	0,1 0	0,1 0	0,3 0	0,1 0	0,1 0	0,1 0	0,2 0	0,2 0	0,8 0	0,7 0	0,5 0	0,7 0	0,0 0	0,1 0	0,1 0	0,1 0	0,2 0	0,2 0	0,8 0	0,1 0	1,0 0	

	373	374	376	380	383	385	387	388	390	393	399	403	405	411	417	418	419	421	422	427	428	429	430	431
B1	0,27	0,64	0,27	0,09	0,09	0,09	0,09	0,27	0,55	0,09	0,64	0,55	0,09	0,82	1,00	1,00	1,00	0,09	0,09	0,18	0,18	0,09	0,09	0,91
B2	0,27	0,64	0,27	0,09	0,18	0,18	0,09	0,27	0,45	0,09	0,55	0,64	0,09	0,27	0,09	0,09	0,09	0,09	0,00	0,91	0,18	0,09	0,09	0,18
B3	0,00	0,50	0,17	0,00	0,17	0,17	0,17	0,17	0,50	0,00	0,33	0,50	0,17	0,17	0,17	0,17	0,17	0,17	0,17	0,17	0,17	0,00	0,00	0,17
G1	0,22	0,33	0,22	0,00	0,89	0,89	0,11	0,33	0,44	0,11	0,67	0,67	0,11	0,22	0,00	0,00	0,00	0,11	0,11	0,11	0,22	0,11	0,11	0,00
G2	0,33	0,50	0,33	0,08	0,17	0,17	0,08	0,25	0,42	0,08	0,50	0,58	0,08	0,25	0,08	0,08	0,17	0,08	0,08	0,17	1,00	1,00	0,08	0,17
G3	0,25	0,67	0,83	0,00	0,17	0,17	1,00	1,00	0,67	0,08	0,42	0,50	1,00	0,25	0,08	0,08	0,17	1,00	0,08	0,08	1,00	0,08	0,08	0,08
R1	0,27	0,36	0,27	0,00	0,18	0,18	0,09	0,27	0,64	0,09	0,45	0,55	0,09	0,27	0,09	0,09	0,18	0,09	1,00	0,00	0,18	0,09	0,09	0,09
R2	0,80	0,50	0,70	0,10	0,20	0,20	0,10	0,10	0,60	0,00	0,60	0,50	0,10	0,10	0,10	0,10	0,20	0,10	0,10	0,20	0,20	0,10	0,00	0,20
R3	0,36	0,55	0,36	0,09	0,00	0,00	0,09	0,27	0,55	0,09	0,36	0,64	0,09	0,27	0,09	0,09	0,18	0,09	0,09	0,18	0,18	0,09	0,09	0,18
W1	0,80	0,50	0,30	0,10	0,20	0,20	0,10	1,00	0,40	0,00	0,60	0,50	0,10	0,90	0,10	0,10	0,20	0,10	0,10	0,20	0,20	0,10	1,00	0,20
W2	0,75	0,50	0,75	1,00	0,13	0,13	0,00	0,13	0,75	0,13	0,63	0,38	0,00	0,25	0,13	0,13	0,25	0,00	0,00	0,88	0,13	0,13	0,00	0,88
Y1	0,67	0,67	0,67	0,11	0,11	0,11	0,11	0,33	0,56	0,11	0,44	0,56	0,11	0,22	0,00	0,00	1,00	0,11	0,11	0,11	0,22	0,11	0,11	0,11
Y2	0,36	0,55	0,36	0,09	0,91	0,91	0,09	0,27	0,36	0,09	0,55	0,55	0,09	0,27	0,09	0,09	0,18	0,09	0,09	0,18	0,18	0,09	0,09	0,18
Y3	0,20	0,50	0,30	0,10	0,20	0,20	0,10	1,00	0,40	1,00	0,40	0,50	0,10	0,90	0,10	0,10	0,20	0,10	0,10	0,20	0,20	0,10	0,00	0,20

Tabelle A.VII: Matrix 2: Angegeben sind die Positionswerte der 82 Positionen bei den 14 Allelen

	7	23	24	42	54	57	83	93	104	14 4	162	17 3	190	201	205	21 1	219	221	230	237	246	247	251	256	259	265	267	269	270
B 1	33, 64	33, 64	41, 63	41, 00	21, 29	0,9 2	6,4 5	10, 64	2,0 9	0, 00	2,8 9	0, 44	31, 22	36, 42	8,5 9	0, 00	29, 43	8,1 5	14,0 0	17,7 9	14, 31	31, 22	39, 00	15,9 5	5,2 0	8,9 0	8,1 5	16,4 6	10, 41
B 2	21, 02	21, 02	26, 02	35, 14	13, 31	10, 12	6,4 5	15, 97	2,0 9	0, 00	2,8 9	0, 44	26, 02	31, 22	8,5 9	0, 00	25, 23	8,1 5	14,0 0	17,7 9	16, 70	36, 42	7,8 0	15,9 5	5,2 0	97, 87	8,1 5	16,4 6	15, 61
B 3	23, 13	23, 13	28, 62	42, 95	14, 64	1,6 9	0,0 0	19, 52	0,0 0	0, 00	0,0 0	0, 00	19, 08	19, 08	0,0 0	0, 00	15, 42	0,0 0	4,80 0	0,00 0	13, 12	28, 62	7,1 5	0,00 0	0,0 0	16, 31	0,0 0	30,1 7	9,5 4
G 1	30, 83	30, 83	38, 16	42, 95	19, 52	1,1 2	7,8 9	45, 54	2,5 6	0, 00	3,5 3	0, 54	25, 44	31, 80	10, 50	0, 00	25, 69	0,0 0	3,20 5	21,7 5	14, 58	31, 80	0,0 0	0,00 0	0,0 0	10, 87	9,9 6	20,1 1	44, 52
G 2	23, 13	23, 13	28, 62	37, 58	14, 64	0,8 4	5,9 1	14, 64	1,9 2	0, 00	2,6 5	0, 81	33, 39	28, 62	7,8 8	0, 00	19, 27	7,4 7	12,8 3	16,3 1	15, 31	33, 39	7,1 5	14,6 2	4,7 7	8,1 6	7,4 7	15,0 8	14, 31
G 3	26, 98	26, 98	33, 39	32, 22	17, 08	0,8 4	5,9 1	48, 79	1,9 2	0, 00	2,6 5	0, 81	23, 85	28, 62	7,8 8	0, 00	26, 98	0,0 0	28,8 3	16,3 1	15, 31	33, 39	3,5 7	0,00 0	0,0 0	8,1 6	7,4 7	181, 01	14, 31
R 1	25, 23	25, 23	31, 22	41, 00	15, 97	0,0 0	6,4 5	15, 97	2,0 9		2,8 9	0, 88	36, 42	31, 22	8,5 9	0, 00	21, 02	0,0 0	2,62 9	17,7 9	14, 31	31, 22	3,9 0	0,00 0	0,0 0	8,1 5	16,4 6	15, 61	
R 2	27, 75	27, 75	34, 34	32, 22	17, 57	1,0 1	7,0 97	46, 84	0,0 0	0, 00	0,0 0	0, 49	34, 34	34, 34	0,0 0	0, 00	23, 13	8,9 6	15,4 0	0,00 0	13, 12	28, 62	8,5 8	17,5 5	5,7 2	9,7 9	89, 62	18,1 0	45, 79
R 3	21, 02	21, 02	26, 02	35, 14	13, 31	0,9 2	6,4 5	10, 64	2,0 9	0, 00	2,8 9	0, 88	31, 22	26, 02	8,5 9	0, 00	16, 82	8,1 5	14,0 0	17,7 9	11, 93	26, 02	7,8 0	15,9 5	5,2 0	8,9 0	8,1 5	16,4 6	5,2 0
W 1	27, 75	27, 75	34, 34	32, 22	17, 57	1,0 1	0,0 0	11, 71	0,0 0	0, 00	0,0 0	0, 49	34, 34	34, 34	0,0 0	0, 00	23, 13	8,9 6	15,4 0	195, 74	13, 12	28, 62	8,5 8	17,5 5	5,7 2	9,7 9	0,0 0	18,1 0	11, 45
W 2	17, 34	17, 34	21, 46	32, 22	10, 98	1,2 7	8,8 7	7,3 2	2,8 8	0, 00	3,9 7	1, 21	35, 77	28, 62	11, 82		17, 34	89, 62	125, 13	0,00 0	16, 40	35, 77	37, 53	175, 47	57, 24	12, 23	11, 20	0,00 0	14, 31
Y 1	30, 83	30, 83	38, 16	35, 79	19, 52	0,0 0	7,8 9	13, 01	2,5 6	0, 00	3,5 3	4, 32	38, 16	44, 52	10, 50	0, 00	10, 28	9,9 6	17,1 1	21,7 5	14, 58	31, 80	4,7 7	19,5 0	6,3 6	0,0 0	9,9 6	20,1 1	12, 72
Y 2	25, 23	25, 23	31, 22	35, 14	15, 97	0,9 2	6,4 5	15, 97	2,0 9	0, 00	2,8 9	0, 88	26, 02	31, 22	8,5 9	0, 00	29, 43	8,1 5	14,0 0	17,7 9	11, 93	26, 02	7,8 0	15,9 5	5,2 0	8,9 0	8,1 5	16,4 6	46, 83
Y 3	18, 50	18, 50	22, 89	32, 22	11, 71	1,0 1	0,0 0	11, 71	2,3 01	0, 00	3,1 78	4, 37	22, 89	34, 34	94, 52	0, 00	23, 13	8,9 6	15,4 0	0,00 0	13, 12	28, 62	8,5 8	17,5 5	5,7 2	9,7 9	0,0 0	18,1 0	11, 45
B 1	33, 64	33, 64	41, 63	41, 00	21, 29	0,9 2	6,4 5	10, 64	2,0 9	0, 00	2,8 9	0, 44	31, 22	36, 42	8,5 9	0, 00	29, 43	8,1 5	14,0 0	17,7 9	14, 31	31, 22	39, 00	15,9 5	5,2 0	8,9 0	8,1 5	16,4 6	10, 41

	273	281	282	287	288	293	295	297	302	303	304	305	307	308	310	312	321	322	323	324	352	353	356	358	359	362	368	370	372
B 1	14,1 4	34, 37	13,1 7	33, 46	29, 08	102, 38	19, 36	4,7 7	2,0 9	2,3 9	26, 43	44, 89	0, 00	2,3 9	7,1 6	25,4 3	34, 37	42, 01	90,4 8	64,0 6	8,5 9	9,34	84, 80		9,8 7		5,3 8	2,6 9	25, 47
B 2	14,1 4	25, 78	13,1 7	62, 14	58, 16	45,5 0	12, 90	23, 86	2,0 9	0,0 0	68, 47	4,0 8		2,3 9	7,1 6	103, 37	34, 37	42, 01	68,5 4	119, 04	8,5 9	9,34	7,7 1			0, 00	5,3 8	29, 61	13, 92
B 3	0,00	78, 77	24,1 5	32, 85	26, 66	62,5 7	47, 32	8,7 5	3,8 4	4,3 7	24, 80	7,4 8	0, 00	4,3 7	26, 24	34,2 5	78, 77	77, 01	58,8 3	71,5 2	0,0 0	17,1 2	14, 13		14, 80		0,0 0	4,9 3	13, 26
G 1	17,2 9	31, 51	16,1 0	31, 40	26, 66	97,3 3	15, 77	2,9 7	2,5 2	2,9 2	32, 31	0,0 0	0, 00	2,9 2	5,8 3	31,0 8	31, 51	30, 81	88,0 0	59,7 5	10, 50	11,4 1	0,0 0			0, 00	3,2 9	3,2 9	11, 67
G 2	12,9 7	78, 77	12,0 8	37, 79	33, 32	41,7 1	11, 83	4,3 7	1,9 2	2,1 9	55, 42	3,7 4	0, 00	2,1 9	6,5 6	23,3 1	31, 51	38, 51	110, 92	58,7 2	7,8 8	8,56	7,0 7			0, 00	7,4 0	2,4 7	13, 92
G 3	12,9 7	23, 63	144, 92	37, 79	33, 32	93,8 5	17, 74	4,3 7	23, 01	2,1 9	24, 23	3,7 4	0, 00	26, 24	6,5 6	23,3 1	39, 39	38, 51	110, 92	58,7 2	7,8 8	102, 72	7,0 7			0, 00	4,9 3	2,4 7	43, 63
R 1	14,1 4	77, 34	13,1 7	51, 07	29, 08	45,5 0	58, 07	2,3 9	2,0 9	26, 24	58, 07	4,0 8	0, 00	2,3 9	7,1 6	103, 37	42, 97	42, 01	68,5 4	119, 04	8,5 9	9,34	7,7 1			0, 00	5,3 8	0,0 0	13, 92
R 2	0,00	28, 36	14,4 9	28, 26	23, 99	50,0 5	14, 19	5,2 5	2,3 0	2,6 2	21, 98	4,4 9	0, 00	2,6 2	20, 99	20,5 5	75, 62	64, 69	60,4 3	59,6 0	0,0 0	10,2 7	8,4 8			0, 00	5,9 2	2,9 6	
R 3	14,1 4	34, 37	13,1 7	55, 20	50, 89	22,7 5	58, 07	23, 86	2,0 9	2,3 9	26, 43	4,0 8	0, 00	2,3 9	9,5 4	25,4 3	60, 15	50, 41	50,0 5	46,6 0	8,5 9	9,34	7,7 1		29, 61		8,0 8	2,6 9	15, 90
W 1	0,00	28, 36	14,4 9	60, 72	55, 98	50,0 5	14, 19	5,2 5	2,3 0	2,6 2	52, 31	4,4 9	0, 00	2,6 2	5,2 5	81,2 9	18, 90	27, 73	103, 17	53,7 8	94, 52	10,2 7	8,4 8		14, 80		5,9 2	2,9 6	
W 2	19,4 5	82, 71	0,00	42, 72	39, 99	31,2 8	8,8 7	6,5 6	0,0 0	0,0 0	18, 60	5,6 1	0, 00	0,0 0	19, 68	12,8 4	47, 26	46, 21	44,1 2	42,4 6	0,0 0	0,00	10, 60			0, 00	22, 21	3,7 0	
Y 1	17,2 9	31, 51	16,1 0	57, 97	53, 32	27,8 1	15, 77	2,9 2	2,5 6	2,9 2	23, 66	0,0 0	0, 00	2,9 2	20, 41	19,6 6	63, 02	51, 34	61,1 7	47,0 2	10, 50	11,4 1	0,0 0			0, 00	23, 03	0,0 0	
Y 2	14,1 4	25, 78	13,1 7	33, 46	29, 08	91,0 1	6,4 5	2,3 9	2,0 9	2,3 9	26, 43	4,0 8	0, 00	2,3 9	7,1 6	25,4 3	25, 78	33, 61	81,2 4	56,4 7	8,5 9	9,34	7,7 1			0, 00	8,0 8	2,6 9	15, 90
Y 3	155, 59	28, 36	14,4 9	60, 72	55, 98	50,0 5	14, 19	5,2 5	2,3 0	2,6 2	21, 98	4,4 9	0, 00	2,6 2	5,2 5	20,5 5	75, 62	64, 69	60,4 3	59,6 0	0,0 0	10,2 7	8,4 8			0, 00	23, 69	2,9 6	27, 17

	373	374	376	380	383	385	387	388	390	393	399	403	405	411	417	418	419	421	422	427	428	429	430	431
B1	0,0 0	18,8 4	53,38	8,90	6,75	8,90	3,61	6,56	5,5 2	2,66	53,0 9	40,4 8	6,75	60,7 3		21,2 8	97,3 0	6,75	10,26	17,7 9	15,64	5,32	2,21	
B2	0,0 0	18,8 4	53,38	8,90	13,4 9	17,7 9	3,61	6,56	4,6 0	2,66	45,5 1	47,2 3	6,75	20,2 4		1,93	8,85	6,75	0,00	88,9 7	15,64	5,32	2,21	0,0 0
B3	0,0 0	14,8 0	32,62	0,00	12,3 7	16,3 1	6,61	4,93	5,0 6	0,00	27,8 1	37,1 1	12,3 7	12,3 7		3,55	16,2 2	12,3 7	18,81	16,3 1	12,37	0,00	0,00	0,0 0
G1	0,0 0	9,87	43,50	0,00	65,9 7	87,0 0	4,41	8,02	4,5 0	3,25	55,6 2	49,4 8	8,25	16,4 9		0,00	0,00	8,25	12,54	10,8 7	19,12	6,51	2,70	0,0 0
G2	0,0 0	14,8 0	65,25	8,16	12,3 7	16,3 1	3,31	6,01	4,2 2	2,44	41,7 2	43,3 0	6,19	18,5 6		1,77	10,5 1	6,19	9,41	16,3 1	101,7 9	58,5 5	2,03	0,0 0
G3	0,0 0	19,7 4	163,1 2	0,00	12,3 7	16,3 1	39,6 8	33,0 0	6,7 5	2,44	34,7 6	37,1 1	74,2 2	18,5 6		1,77	10,5 1	74,2 2	9,41	8,16	80,11	4,88	2,03	0,0 0
R1	0,0 0	10,7 7	53,38	0,00	13,4 9	17,7 9	3,61	6,56	6,4 4	2,66	37,9 2	40,4 8	6,75	20,2 4		1,93	11,4 7	6,75	112,8 7	0,00	15,64	5,32	2,21	0,0 0
R2		14,8 0	137,0 2	9,79	14,8 4	19,5 7	3,97	2,96	6,0 7	0,00	50,0 6	37,1 1	7,42	7,42		2,13	12,6 1	7,42	11,29	19,5 7	17,21	5,85	0,00	0,0 0
R3	0,0 0	16,1 5	71,18	8,90	0,00	0,00	3,61	6,56	5,5 2	2,66	30,3 4	47,2 3	6,75	20,2 4		1,93	11,4 7	6,75	10,26	17,7 9	15,64	5,32	2,21	0,0 0
W1		14,8 0	58,72	9,79	14,8 4	19,5 7	3,97	24,1 4	4,0 5	0,00	50,0 6	37,1 1	7,42	66,8 0		2,13	12,6 1	7,42	11,29	19,5 7	17,21	5,85	24,3 2	0,0 0
W2		14,8 0	146,8 1	97,8 7	9,28	12,2 3	0,00	2,66	7,5 9	3,66	52,1 5	27,8 3	0,00	18,5 6		2,66	15,7 7	0,00	0,00	85,6 4	12,23	7,32	0,00	
Y1		19,7 4	130,5 0	10,8 7	8,25 7	10,8 7	4,41	8,02	5,6 2	3,25	37,0 8	41,2 3	8,25	16,4 9		0,00	28,8 3	8,25	12,54	10,8 7	19,12	6,51	2,70	0,0 0
Y2	0,0 0	16,1 5	71,18	8,90	67,4 7	88,9 7	3,61	6,56	3,6 8	2,66	45,5 1	40,4 8	6,75	20,2 4		1,93	11,4 7	6,75	10,26	17,7 9	15,64	5,32	2,21	0,0 0
Y3	0,0 0	14,8 0	58,72	9,79	14,8 4	19,5 7	3,97	24,1 4	4,0 5	29,2 8	33,3 7	37,1 1	7,42	66,8 0		2,13	12,6 1	7,42	11,29	19,5 7	17,21	5,85	0,00	0,0 0
B1	0,0 0	18,8 4	53,38	8,90	6,75	8,90	3,61	6,56	5,5 2	2,66	53,0 9	40,4 8	6,75	60,7 3		21,2 8	97,3 0	6,75	10,26	17,7 9	15,64	5,32	2,21	

Tabelle A.VIII: Matrix 3 Angegeben sind die Positionswerte der 72 Positionen bei den 14 Allelen

	7	23	24	42	54	57	83	93	104	162	173	190	201	205	219	221	230	237	246	247	251	256	259	265	267	269	270
B1	33,6 4	33,6 4	41,6 3	41,0 0	21,2 9	0,92	6,45	10,6 4	2,09	2,89	0,4 4	31,2 2	36,4 2	8,59	29,4 3	8,15	14,0 0	17,7 9	14,3 1	31,2 2	39,0 0	15,9 5	5,20	8,90	8,15	16,4 6	10,4 1
B2	21,0 2	21,0 2	26,0 2	35,1 4	13,3 1	10,1 2	6,45	15,9 7	2,09	2,89	0,4 4	26,0 2	31,2 2	8,59	25,2 3	8,15	14,0 0	17,7 9	16,7 0	36,4 2	7,80	15,9 5	5,20	97,8 7	8,15	16,4 6	15,6 1
B3	23,1 3	23,1 3	28,6 2	42,9 5	14,6 4	1,69	0,00	19,5 2	0,00	0,00	0,0 0	19,0 8	19,0 8	0,00	15,4 2	0,00	4,80	0,00	13,1 2	28,6 2	7,15	0,00	0,00	16,3 1	0,00	30,1 7	9,54
G1	30,8 3	30,8 3	38,1 6	42,9 5	19,5 2	1,12	7,89	45,5 4	2,56	3,53	0,5 4	25,4 4	31,8 0	10,5 0	25,6 9	0,00	3,20	21,7 5	14,5 8	31,8 0	0,00	0,00	0,00	10,8 7	9,96	20,1 1	44,5 2
G2	23,1 3	23,1 3	28,6 2	37,5 8	14,6 4	0,84	5,91	14,6 4	1,92	2,65	0,8 1	33,3 9	28,6 2	7,88	19,2 7	7,47	12,8 3	16,3 1	15,3 1	33,3 9	7,15	14,6 2	4,77	8,16	7,47	15,0 8	14,3 1
G3	26,9 8	26,9 8	33,3 9	32,2 2	17,0 8	0,84	5,91	48,7 9	1,92	2,65	0,8 1	23,8 5	28,6 2	7,88	26,9 8	0,00	28,8 3	16,3 1	15,3 1	33,3 9	3,57	0,00	0,00	8,16	7,47	181, 01	14,3 1
R1	25,2 3	25,2 3	31,2 2	41,0 0	15,9 7	0,00	6,45	15,9 7	2,09	2,89	0,8 8	36,4 2	31,2 2	8,59	21,0 2	0,00	2,62	17,7 9	14,3 1	31,2 2	3,90	0,00	0,00	0,00	8,15	16,4 6	15,6 1
R2	27,7 5	27,7 5	34,3 4	32,2 2	17,5 7	1,01	70,9 7	46,8 4	0,00	0,00	0,4 9	34,3 4	34,3 4	0,00	23,1 3	8,96	15,4 0	0,00	13,1 2	28,6 2	8,58	17,5 5	5,72	9,79	89,6 2	18,1 0	45,7 9
R3	21,0 2	21,0 2	26,0 2	35,1 4	13,3 1	0,92	6,45	10,6 4	2,09	2,89	0,8 8	31,2 2	26,0 2	8,59	16,8 2	8,15	14,0 0	17,7 9	11,9 3	26,0 2	7,80	15,9 5	5,20	8,90	8,15	16,4 6	5,20
W1	27,7 5	27,7 5	34,3 4	32,2 2	17,5 7	1,01	0,00	11,7 1	0,00	0,00	0,4 9	34,3 4	34,3 4	0,00	23,1 3	8,96	15,4 0	19,5 74	13,1 2	28,6 2	8,58	17,5 5	5,72	9,79	0,00	18,1 0	11,4 5
W2	17,3 4	17,3 4	21,4 6	32,2 2	10,9 8	1,27	8,87	7,32	2,88	3,97	1,2 1	35,7 7	28,6 2	11,8 2	17,3 4	89,6 2	12,5 13	0,00	16,4 0	35,7 7	37,5 3	17,5 47	57,2 4	12,2 3	11,2 0	0,00	14,3 1
Y1	30,8 3	30,8 3	38,1 6	35,7 9	19,5 2	0,00	7,89	13,0 1	2,56	3,53	4,3 2	38,1 6	44,5 2	10,5 0	10,2 8	9,96	17,1 1	21,7 5	14,5 8	31,8 0	4,77	19,5 0	6,36	0,00	9,96	20,1 1	12,7 2
Y2	25,2 3	25,2 3	31,2 2	35,1 4	15,9 7	0,92	6,45	15,9 7	2,09	2,89	0,8 8	26,0 2	31,2 2	8,59	29,4 3	8,15	14,0 0	17,7 9	11,9 3	26,0 2	7,80	15,9 5	5,20	8,90	8,15	16,4 6	46,8 3
Y3	18,5 0	18,5 0	22,8 9	32,2 2	11,7 1	1,01	0,00	11,7 1	20,7 1	28,6 1	4,3 7	28,6 2	34,3 4	85,0 7	18,5 0	8,96	15,4 0	0,00	10,5 0	22,8 9	4,29	17,5 5	5,72	9,79	0,00	18,1 0	11,4 5
B1	33,6 4	33,6 4	41,6 3	41,0 0	21,2 9	0,92	6,45	10,6 4	2,09	2,89	0,4 4	31,2 2	36,4 2	8,59	29,4 3	8,15	14,0 0	17,7 9	14,3 1	31,2 2	39,0 0	15,9 5	5,20	8,90	8,15	16,4 6	10,4 1

	273	281	282	287	288	293	295	297	302	303	304	305	308	310	312	321	322	323	324	352	353	356	368	370
B1	14,14	34,3 7	13,17	33,4 6	29,0 8	102,3 8	19,3 6	4,77	2,09	2,39	26,4 3	44,8 9	2,39	7,16	25,43	34,3 7	42,0 1	90,48	64,06	8,59	9,34	84,8 0	5,38	2,69
B2	14,14	25,7 8	13,17	62,1 4	58,1 6	45,50	12,9 0	23,8 6	2,09	0,00	68,4 7	4,08	2,39	7,16	103,3 7	34,3 7	42,0 1	68,54	119,0 4	8,59	9,34	7,71	5,38	29,6 1
B3	0,00	78,7 7	24,15	32,8 5	26,6 6	62,57	47,3 2	8,75	3,84	4,37	24,8 0	7,48	4,37	26,2 4	34,25	78,7 7	77,0 1	58,83	71,52	0,00	17,12	14,1 3	0,00	4,93
G1	17,29	31,5 1	16,10	31,4 0	26,6 6	97,33	15,7 7	2,92	2,56	2,92	32,3 1	0,00	2,92	5,83	31,08	31,5 1	30,8 1	88,00	59,75	10,5 0	11,41	0,00	3,29	3,29
G2	12,97	78,7 7	12,08	37,7 9	33,3 2	41,71	11,8 3	4,37	1,92	2,19	55,4 2	3,74	2,19	6,56	23,31	31,5 1	38,5 1	110,9 2	58,72	7,88	8,56	7,07	7,40	2,47
G3	12,97	23,6 3	144,9 2	37,7 9	33,3 2	93,85	17,7 4	4,37	23,0 1	2,19	24,2 3	3,74	4	6,56	23,31	39,3 9	38,5 1	110,9 2	58,72	7,88	102,7 2	7,07	4,93	2,47
R1	14,14	77,3 4	13,17	51,0 7	29,0 8	45,50	58,0 7	2,39	2,09	26,2 4	58,0 7	4,08	2,39	7,16	103,3 7	42,9 7	42,0 1	68,54	119,0 4	8,59	9,34	7,71	5,38	0,00
R2	0,00	28,3 6	14,49	28,2 6	23,9 9	50,05	14,1 9	5,25	2,30	2,62	21,9 8	4,49	2,62	20,9 9	20,55	75,6 2	64,6 9	60,43	59,60	0,00	10,27	8,48	5,92	2,96
R3	14,14	34,3 7	13,17	55,2 0	50,8 9	22,75	58,0 7	23,8 6	2,09	2,39	26,4 3	4,08	2,39	9,54	25,43	60,1 5	50,4 1	50,05	46,60	8,59	9,34	7,71	8,08	2,69
W1	0,00	28,3 6	14,49	60,7 2	55,9 8	50,05	14,1 9	5,25	2,30	2,62	52,3 1	4,49	2,62	5,25	81,29	18,9 0	27,7 3	103,1 7	53,78	94,5 2	10,27	8,48	5,92	2,96
W2	19,45	82,7 1	0,00	42,7 2	39,9 9	31,28	8,87	6,56	0,00	0,00	18,6 0	5,61	0,00	19,6 8	12,84	47,2 6	46,2 1	44,12	42,46	0,00	0,00	10,6 0	22,2 1	3,70
Y1	17,29	31,5 1	16,10	57,9 7	53,3 2	27,81	15,7 7	2,92	2,56	2,92	23,6 6	0,00	2,92	20,4 1	19,66	63,0 2	51,3 4	61,17	47,02	10,5 0	11,41	0,00	23,0 3	0,00
Y2	14,14	25,7 8	13,17	33,4 6	29,0 8	91,01	6,45	2,39	2,09	2,39	26,4 3	4,08	2,39	7,16	25,43	25,7 8	33,6 1	81,24	56,47	8,59	9,34	7,71	8,08	2,69
Y3	140,0 3	37,8 1	14,49	60,7 2	55,9 8	37,54	14,1 9	5,25	2,30	2,62	29,0 8	0,00	2,62	7,87	20,55	75,6 2	64,6 9	65,22	59,60	0,00	10,27	0,00	23,6 9	2,96

	374	376	380	383	385	387	388	390	393	399	403	405	411	418	419	421	422	427	428	429	430
B1	18,84	53,38	8,90	6,75	8,90	3,61	6,56	5,52	2,66	53,09	40,48	6,75	60,73	21,28	97,30	6,75	10,26	17,79	15,64	5,32	2,21
B2	18,84	53,38	8,90	13,49	17,79	3,61	6,56	4,60	2,66	45,51	47,23	6,75	20,24	1,93	8,85	6,75	0,00	88,97	15,64	5,32	2,21
B3	14,80	32,62	0,00	12,37	16,31	6,61	4,93	5,06	0,00	27,81	37,11	12,37	12,37	3,55	16,22	12,37	18,81	16,31	12,37	0,00	0,00
G1	9,87	43,50	0,00	65,97	87,00	4,41	8,02	4,50	3,25	55,62	49,48	8,25	16,49	0,00	0,00	8,25	12,54	10,87	19,12	6,51	2,70
G2	14,80	65,25	8,16	12,37	16,31	3,31	6,01	4,22	2,44	41,72	43,30	6,19	18,56	1,77	10,51	6,19	9,41	16,31	101,79	58,55	2,03
G3	19,74	163,12	0,00	12,37	16,31	39,68	33,00	6,75	2,44	34,76	37,11	74,22	18,56	1,77	10,51	74,22	9,41	8,16	80,11	4,88	2,03
R1	10,77	53,38	0,00	13,49	17,79	3,61	6,56	6,44	2,66	37,92	40,48	6,75	20,24	1,93	11,47	6,75	112,87	0,00	15,64	5,32	2,21
R2	14,80	137,02	9,79	14,84	19,57	3,97	2,96	6,07	0,00	50,06	37,11	7,42	7,42	2,13	12,61	7,42	11,29	19,57	17,21	5,85	0,00
R3	16,15	71,18	8,90	0,00	0,00	3,61	6,56	5,52	2,66	30,34	47,23	6,75	20,24	1,93	11,47	6,75	10,26	17,79	15,64	5,32	2,21
W1	14,80	58,72	9,79	14,84	19,57	3,97	24,14	4,05	0,00	50,06	37,11	7,42	66,80	2,13	12,61	7,42	11,29	19,57	17,21	5,85	24,32
W2	14,80	146,81	97,87	9,28	12,23	0,00	2,66	7,59	3,66	52,15	27,83	0,00	18,56	2,66	15,77	0,00	0,00	85,64	12,23	7,32	0,00
Y1	19,74	130,50	10,87	8,25	10,87	4,41	8,02	5,62	3,25	37,08	41,23	8,25	16,49	0,00	28,83	8,25	12,54	10,87	19,12	6,51	2,70
Y2	16,15	71,18	8,90	67,47	88,97	3,61	6,56	3,68	2,66	45,51	40,48	6,75	20,24	1,93	11,47	6,75	10,26	17,79	15,64	5,32	2,21
Y3	14,80	78,30	9,79	14,84	19,57	3,97	22,02	4,05	26,35	25,03	37,11	7,42	66,80	0,00	5,77	7,42	11,29	19,57	27,00	11,71	0,00
B1	18,84	53,38	8,90	6,75	8,90	3,61	6,56	5,52	2,66	53,09	40,48	6,75	60,73	21,28	97,30	6,75	10,26	17,79	15,64	5,32	2,21

Tabelle A. IX: Koordinatenwerte der MDS der Allele von Matrix 1 Allele nach Spearman'scher Rangkorrelation

	x	y
B1	-1,19	0,67
B2	-0,35	0,75
B3	0,23	0,84
G1	0,77	-0,04
G2	-0,02	-0,06
G3	1,78	-0,20
R1	0,74	0,32
R2	0,12	0,41
R3	-0,28	0,07
W1	0,48	-0,72
W2	-1,85	-0,41
Y1	-0,01	-0,55
Y2	0,22	-0,15
Y3	-0,63	-0,94

Tabelle A. X: Koordinatenwerte der MDS der Allele von Matrix 1 Allele nach Bildung der Euklidischen Distanzen

	x	y
B1	-0,71	1,40
B2	0,05	-1,20
B3	-0,01	0,18
G1	0,18	0,20
G2	0,01	0,08
G3	1,39	1,05
R1	-0,65	0,24
R2	0,35	0,13
R3	-0,10	-0,05
W1	-1,24	0,00
W2	1,32	-1,00
Y1	0,13	-0,03
Y2	0,20	0,04
Y3	-0,93	-1,05

Tabelle A.XI: MDS der Positionen von Matrix 1 nach Bildung der Euklidischen Distanzen: Gruppen von Positionen mit gleichen Koordinatenwerten sind durch Grauschattierungen hervorgehoben. Positionen mit ähnlichen Koordinatenwerten (gestichelter Kreis in Abbildung III.10) sind kursiv gedruckt.

Position	x	y	Position	x	y
7	-1,08	-0,06	104	1,07	0,24
23	-1,08	-0,06	162	1,07	0,24
24	-1,08	-0,06	205	1,07	0,24
54	-1,08	-0,06	273	1,07	0,24
246	-0,94	-0,01	393	1,07	0,24
247	-0,94	-0,01	310	-0,80	0,84
42	-1,07	-0,02	304	-1,14	-0,56
403	-1,02	-0,08	372	-2,00	0,14
374	-0,89	0,06	427	0,32	0,72
190	-0,93	0,03	144	0,77	-0,72
201	-1,07	0,03	303	0,77	-0,72
390	-0,85	0,10	422	0,77	-0,72
399	-0,84	-0,02	373	-0,50	0,50
211	0,94	0,68	93	-0,29	-0,53
221	0,94	0,68	383	0,12	-0,93
256	0,94	0,68	385	0,12	-0,93
259	0,94	0,68	237	0,97	-0,55
380	0,94	0,68	352	0,97	-0,55
323	-1,73	-0,40	430	0,97	-0,55
287	-1,44	0,02	281	-0,91	0,60
322	-0,91	0,31	368	-0,07	0,59
324	-1,83	-0,05	362	-0,45	0,00
269	0,66	-0,44	295	-0,22	-0,18
282	0,66	-0,44	419	-0,09	1,00
302	0,66	-0,44	83	1,10	-0,15
308	0,66	-0,44	267	1,10	-0,15
353	0,66	-0,44	428	-0,13	-0,87
387	0,66	-0,44	270	-0,38	-0,49
405	0,66	-0,44	376	-0,41	0,18
421	0,66	-0,44	312	-0,98	-0,86
219	-0,63	-0,18	297	-0,10	0,75
251	0,25	0,66	359	-1,17	0,50
431	0,25	0,66	230	0,13	0,55
305	0,75	0,20	173	0,23	0,92
356	0,75	0,20	388	-0,63	-0,97
358	0,75	0,20	411	-0,26	0,18
417	0,75	0,20			
418	0,75	0,20			
321	-1,00	0,54			
293	-0,79	-0,44			
57	0,89	-0,03			
265	0,89	-0,03			
307	0,89	-0,03			
370	0,89	-0,03			
288	-0,74	0,24			
429	0,92	-0,18			

Tabelle A. XII: Koordinatenwerte der MDS der Allele von Matrix 3 Allele nach Spearman'scher Rangkorrelation

B1	-0,23	-1,01
B2	-0,50	-0,42
B3	0,71	-0,42
G1	0,66	0,37
G2	-0,12	0,00
G3	1,60	0,33
R1	0,91	-0,07
R2	0,17	0,86
R3	-0,48	-0,07
W1	0,29	-0,85
W2	-2,01	-0,05
Y1	-0,23	0,17
Y2	0,05	0,13
Y3	-0,83	1,03

Tabelle A:XIII: Koordinatenwerte der MDS der Allele von Matrix 3 Allele nach Bildung der Euklidischen Distanzen

B1	0,11	0,15
B2	-0,19	-0,38
B3	-0,11	-0,04
G1	-0,01	0,00
G2	-0,06	-0,03
G3	1,99	0,95
R1	-0,02	-0,32
R2	-0,25	0,16
R3	-0,01	-0,09
W1	1,18	-1,26
W2	-2,31	0,17
Y1	-0,15	0,07
Y2	-0,04	0,06
Y3	-0,13	0,56

Tabelle A.XIV: MDS der Positionen von Matrix 3 nach Bildung der Euklidischen Distanzen: Gruppen von Positionen mit gleichen Koordinatenwerten sind durch Grauschattierungen hervorgehoben. Koordinatenwerte wurden bis zu einem Unterschied von maximal 0,1 als ähnlich bewertet.

Aminosäure-	x	y	Aminosäure-	x	y
393	-0,68	-0,03	288	0,63	-0,2
162	-0,68	-0,03	405	-0,42	0,38
104	-0,72	-0,02	421	-0,42	0,38
173	-0,77	0,02	399	0,72	-0,05
57	-0,76	0,03	287	0,9	-0,1
418	-0,71	0,01	322	1,03	-0,24
370	-0,64	0	295	0,09	0,18
310	-0,37	-0,07	83	-0,73	-0,21
430	-0,73	0,08	383	-0,06	0,39
303	-0,68	0,08	221	-0,61	-0,63
302	-0,66	0,11	380	-0,6	-0,73
308	-0,65	0,13	293	1,79	0,54
297	-0,52	-0,03	411	0,25	-0,32
390	-0,58	0,01	428	0,31	0,76
388	-0,41	0,12	321	1,16	-0,41
251	-0,45	-0,22	427	0,13	-0,81
246	-0,28	-0,02	205	-0,44	-0,48
374	-0,24	-0,03	353	-0,24	0,67
54	-0,22	0	281	1,07	-0,59
305	-0,6	-0,11	356	-0,54	-0,4
387	-0,56	0,2	323	2,72	0,59
368	-0,46	-0,11	324	2,19	-0,02
270	-0,09	0,1	267	-0,86	-0,29
93	-0,08	0,23	312	1,17	0,47
7	0,05	0	385	0,34	0,59
23	0,05	0	265	-0,25	-0,59
219	-0,07	-0,01	419	-0,11	-0,49
190	0,22	-0,08	282	0,1	1,22
201	0,26	-0,06	269	0,51	1,64
247	0,21	-0,03	230	-0,3	-1,02
24	0,25	0,01	376	3,68	-0,76
403	0,61	0	352	-0,74	0,53
259	-0,61	-0,32	256	-0,35	-1,79
429	-0,5	0,11	422	-0,45	0,8
304	0,54	0,1	273	0,06	-1,25
42	0,44	-0,02	237	-0,65	2,05

Tabelle A.XV.: Spearmansch'e Rangkorrelationskoeffizienten der Allele aus Matrix 1. Der kleinste und der größte Korrelationskoeffizient sind fett gedruckt.

	B1	B2	B3	G1	G2	G3	R1	R2	R3	W1	W2	Y1	Y2	Y3
B1														
B2	0,42													
B3	0,62	0,66												
G1	0,31	0,63	0,67											
G2	0,57	0,67	0,68	0,77										
G3	0,25	0,32	0,56	0,59	0,50									
R1	0,47	0,38	0,71	0,73	0,72	0,45								
R2	0,50	0,58	0,64	0,67	0,73	0,43	0,59							
R3	0,63	0,68	0,77	0,57	0,79	0,46	0,68	0,67						
W1	0,54	0,59	0,56	0,63	0,70	0,40	0,60	0,47	0,69					
W2	0,42	0,44	0,27	0,18	0,52	-0,04	0,16	0,53	0,54	0,32				
Y1	0,43	0,45	0,54	0,65	0,79	0,52	0,71	0,70	0,79	0,64	0,53			
Y2	0,56	0,65	0,63	0,86	0,84	0,50	0,68	0,76	0,63	0,70	0,50	0,76		
Y3	0,38	0,45	0,39	0,56	0,61	0,32	0,51	0,34	0,61	0,34	0,42	0,66	0,63	

Tabelle A.XVI: Euklidischen Distanzen der Allele aus Matrix 1. Die kleinste und die größte Distanz sind fett gedruckt.

	B1	B2	B3	G1	G2	G3	R1	R2	R3	W1	W2	Y1	Y2
B1													
B2	3,64												
B3	2,95	2,70											
G1	3,35	2,83	2,25										
G2	3,10	2,72	2,16	2,03									
G3	4,06	3,97	3,37	3,27	3,26								
R1	3,46	3,12	2,25	2,51	2,40	3,79							
R2	3,31	3,08	2,36	2,20	2,39	3,47	2,84						
R3	2,94	2,54	1,51	2,43	2,08	3,44	2,36	2,37					
W1	3,23	3,03	2,93	2,75	2,67	3,70	2,88	3,22	2,48				
W2	3,68	3,59	3,34	3,63	3,14	4,38	3,81	3,13	3,08	3,78			
Y1	3,18	3,11	2,35	2,46	2,17	3,41	2,63	2,17	2,04	2,81	2,95		
Y2	3,02	2,78	2,26	0,92	1,95	3,31	2,57	2,12	2,31	2,66	3,19	2,18	
Y3	3,68	3,51	3,25	3,09	3,00	3,88	3,34	3,41	2,87	3,47	3,67	2,67	2,90

TabelleA. XVII: Spearmansch'e Rangkorrelationskoeffizienten der Allele aus Matrix3. Der kleinste und der größte Korrelationskoeffizient sind fett gedruckt.

	B1	B2	B3	G1	G2	G3	R1	R2	R3	W1	W2	Y1	Y2
B2	0,668												
B3	0,714	0,717											
G1	0,540	0,732	0,729										
G2	0,748	0,808	0,708	0,802									
G3	0,502	0,531	0,728	0,712	0,627								
R1	0,677	0,575	0,784	0,820	0,766	0,654							
R2	0,649	0,685	0,741	0,707	0,745	0,572	0,649						
R3	0,759	0,802	0,752	0,626	0,822	0,595	0,730	0,683					
W1	0,696	0,747	0,656	0,694	0,794	0,571	0,689	0,559	0,739				
W2	0,586	0,614	0,378	0,311	0,620	0,108	0,283	0,609	0,615	0,446			
Y1	0,705	0,701	0,701	0,721	0,851	0,630	0,773	0,735	0,860	0,751	0,627		
Y2	0,745	0,812	0,704	0,869	0,889	0,617	0,747	0,782	0,710	0,795	0,612	0,825	
Y3	0,499	0,614	0,532	0,652	0,713	0,498	0,600	0,461	0,656	0,498	0,584	0,715	0,680

TabelleA.XVIII.: Euklidischen Distanzen der Allele aus Matrix3. Die kleinste und die größte Distanz sind fett gedruckt.

	B1	B2	B3	G1	G2	G3	R1	R2	R3	W1	W2	Y1	Y2
B2	223,83												
B3	173,25	187,89											
G1	189,58	202,84	154,77										
G2	191,85	203,22	157,96	167,00									
G3	322,34	341,93	301,90	303,24	297,25								
R1	220,70	198,05	156,13	195,50	192,26	339,31							
R2	215,78	228,62	181,40	188,43	196,22	299,56	226,60						
R3	178,33	179,64	107,90	172,07	149,04	305,55	172,42	160,61					
W1	255,98	252,07	267,49	244,35	246,90	362,73	258,48	294,64	236,77				
W2	313,42	311,43	310,99	332,09	296,48	410,36	345,04	282,81	274,05	367,37			
Y1	184,03	202,59	149,65	176,27	153,99	282,44	194,64	134,40	90,44	240,00	258,90		
Y2	175,70	199,41	159,24	58,72	164,50	301,08	202,27	175,23	161,67	243,38	297,66	154,44	
Y3	235,06	237,83	207,70	220,46	209,57	334,40	239,45	236,35	176,44	298,72	309,57	176,87	211,43

Tabelle A.XXI: Spearman'sche Rangkorrelation der Aminosäurepositionen aus Matrix 3

	7	23	24	42	54	57	83	93	104	162	173	190	201	205	219	221	230	237	246	247	251	256	259	265	267	269	270	273	281	282	287	288	293	295	297	302			
23	1																																						
24	1	1																																					
42	0,32	0,32	0,32																																				
54	1,00	1,00	1,00	0,32																																			
57	-0,36	-0,36	-0,36	-0,09	-0,36																																		
83	0,22	0,22	0,22	0,01	0,22	0,00																																	
93	0,30	0,30	0,30	0,15	0,30	0,02	0,02																																
104	-0,24	-0,24	-0,24	0,01	-0,24	0,00	0,33	-0,44																															
162	-0,24	-0,24	-0,24	0,01	-0,24	0,00	0,33	-0,44	1,00																														
173	-0,32	-0,32	-0,32	-0,33	-0,32	-0,45	0,17	-0,36	0,69	0,69																													
190	0,09	0,09	0,09	-0,20	0,09	-0,43	0,42	-0,48	0,17	0,17	0,44																												
201	0,62	0,62	0,62	-0,10	0,62	-0,18	0,24	-0,13	0,24	0,24	0,09	0,41																											
205	-0,24	-0,24	-0,24	0,01	-0,24	0,00	0,33	-0,44	1,00	1,00	0,69	0,17	0,24																										
219	0,42	0,42	0,42	-0,04	0,42	0,00	0,04	0,33	-0,15	-0,15	-0,37	-0,37	0,23	-0,15																									
221	-0,14	-0,14	-0,14	-0,61	-0,14	0,12	0,32	-0,60	0,32	0,32	0,40	0,60	0,47	0,32	-0,30																								
230	-0,09	-0,09	-0,09	-0,82	-0,09	-0,01	0,16	-0,25	0,16	0,16	0,37	0,26	0,23	0,16	-0,11	0,72																							
237	0,56	0,56	0,56	0,28	0,56	-0,27	0,07	-0,10	0,07	0,07	-0,05	0,18	0,41	0,07	0,24	-0,05	-0,20																						
246	0,01	0,01	0,01	0,09	0,01	0,06	0,31	0,12	0,09	0,09	-0,16	0,09	-0,13	0,09	0,09	-0,10	0,13	0,06	1,00																				
247	0,01	0,01	0,01	0,09	0,01	0,06	0,31	0,12	0,09	0,09	-0,16	0,09	-0,13	0,09	0,09	-0,10	0,13	0,06	1,00																				
251	-0,05	-0,05	-0,05	-0,30	-0,05	0,30	0,21	-0,54	-0,20	-0,20	-0,28	0,30	0,13	-0,20	0,09	0,58	0,30	-0,10	-0,06	-0,06																			
256	-0,14	-0,14	-0,14	-0,61	-0,14	0,12	0,32	-0,60	0,32	0,32	0,40	0,60	0,47	0,32	-0,30	1,00	0,72	-0,05	-0,10	-0,10	0,58																		
259	-0,14	-0,14	-0,14	-0,61	-0,14	0,12	0,32	-0,60	0,32	0,32	0,40	0,60	0,47	0,32	-0,30	1,00	0,72	-0,05	-0,10	-0,10	0,58	1,00																	
265	-0,36	-0,36	-0,36	-0,09	-0,36	1,00	0,00	0,02	0,00	0,00	-0,45	-0,43	-0,18	0,00	0,00	0,12	-0,01	-0,27	0,06	0,06	0,30	0,12	0,12																
267	0,22	0,22	0,22	0,01	0,22	0,00	1,00	0,02	0,33	0,33	0,17	0,42	0,24	0,33	0,04	0,32	0,16	0,07	0,31	0,31	0,21	0,32	0,32	0,00															
269	0,44	0,44	0,44	0,06	0,44	0,02	-0,23	0,60	-0,23	-0,23	-0,19	-0,42	0,16	-0,23	-0,03	-0,27	0,10	0,09	-0,17	-0,17	-0,53	-0,27	-0,27	0,02	-0,23														
270	0,14	0,14	0,14	-0,08	0,14	0,00	0,51	0,56	0,04	0,04	0,05	0,03	0,14	0,04	0,50	-0,09	-0,10	0,04	0,26	0,26	-0,17	-0,09	-0,09	0,00	0,51	-0,12													
273	-0,24	-0,24	-0,24	0,01	-0,24	0,00	0,33	-0,44	1,00	1,00	0,69	0,17	0,24	1,00	-0,15	0,32	0,16	0,07	0,09	0,09	-0,20	0,32	0,32	0,00	0,33	-0,23	0,04												
281	-0,41	-0,41	-0,41	0,33	-0,41	0,05	-0,06	-0,48	0,24	0,24	0,20	0,29	-0,30	0,24	-0,62	0,03	-0,25	-0,40	0,00	0,00	0,07	0,03	0,03	0,05	-0,06	-0,39	-0,40	0,24											
282	0,44	0,44	0,44	0,06	0,44	0,02	-0,23	0,60	-0,23	-0,23	-0,19	-0,42	0,16	-0,23	-0,03	-0,27	0,10	0,09	-0,17	-0,17	-0,53	-0,27	-0,27	0,02	-0,23	1,00	-0,12	-0,23	-0,39										
287	-0,40	-0,40	-0,40	-0,36	-0,40	-0,05	-0,32	-0,49	0,31	0,31	0,36	0,30	0,13	0,31	-0,30	0,38	0,28	0,28	0,08	0,08	0,00	0,38	0,38	-0,05	-0,32	-0,20	-0,31	0,31	-0,05	-0,20									
288	-0,41	-0,41	-0,41	-0,46	-0,41	0,03	-0,30	-0,53	0,33	0,33	0,34	0,21	0,12	0,33	-0,25	0,47	0,43	0,24	0,14	0,14	0,08	0,47	0,47	0,03	-0,30	-0,19	-0,33	0,33	-0,11	-0,19	0,96								
293	0,59	0,59	0,59	0,29	0,59	0,13	-0,11	0,52	-0,35	-0,35	-0,63	-0,56	0,14	-0,35	0,79	-0,53	-0,28	0,12	0,01	0,01	-0,05	-0,53	-0,53	0,13	-0,11	0,36	0,27	-0,35	-0,40	0,36	-0,61	-0,58							
295	0,27	0,27	0,27	0,44	-0,27	-0,29	-0,14	0,09	-0,14	-0,14	-0,16	-0,07	-0,10	-0,14	-0,21	-0,48	-0,35	0,15	-0,22	-0,22	-0,31	-0,48	-0,48	-0,29	-0,14	0,41	-0,51	-0,14	0,14	0,41	-0,08	-0,24	0,09						
297	-0,54	-0,54	-0,54	-0,29	-0,54	0,67	-0,15	-0,30	-0,15	-0,15	-0,32	-0,18	-0,34	-0,15	-0,35	0,27	0,21	-0,36	-0,01	-0,01	0,46	0,27	0,27	0,67	-0,15	-0,11	-0,51	0,16	-0,11	0,28	0,35	-0,33	0,03						
302	0,44	0,44	0,44	0,06	0,44	0,02	-0,23	0,60	-0,23	-0,23	-0,19	-0,42	0,16	-0,23	-0,03	-0,27	0,10	0,09	-0,17	-0,17	-0,53	-0,27	-0,27	0,02	-0,23	1,00	-0,12	-0,23	-0,39	1,00	-0,20	-0,19	0,36	0,41	-0,11				
303	0,41	0,41	0,41	0,47	0,41	-0,18	-0,07	0,26	-0,07	-0,07	0,00	0,12	0,28	-0,07	-0,28	-0,23	-0,44	0,22	-0,47	-0,47	-0,42	-0,23	-0,23	-0,18	-0,07	0,52	-0,08	-0,07	0,17	0,52	-0,21	-0,40	0,11	0,55	-0,36	0,52			
304	-0,08	-0,08	-0,08	0,31	-0,08	0,00	-0,41	0,02	-0,03	-0,03	-0,20	-0,10	0,01	-0,03	0,25	-0,41	-0,64	0,42	0,06	0,06	-0,25	-0,41	-0,41	0,00	-0,41	-0,25	0,10	-0,03	-0,04	-0,25	0,40	0,29	0,06	0,01	-0,10	-0,25			
305	-0,01	-0,01	-0,01	0,01	-0,01	0,36	0,02	-0,24	-0,45	-0,45	-0,56	0,05	-0,17	-0,45	0,09	0,11	0,01	0,36	0,02	-0,28	-0,24	-0,45	0,11	0,11	0,36	0,02	-0,28	-0,24	-0,45	0,21	-0,28	-0,26	-0,27	0,25	0,08	0,42	-0,28		
308	0,44	0,44	0,44	0,06	0,44	0,02	-0,23	0,60	-0,23	-0,23	-0,19	-0,42	0,16	-0,23	-0,03	-0,27	0,10	0,09	-0,17	-0,17	-0,53	-0,27	-0,27	0,02	-0,23	1,00	-0,12	-0,23	-0,39	1,00	-0,20	-0,19	0,36	0,41	-0,11	1,00			
310	-0,28	-0,28	-0,28	-0,03	-0,28	0,17	0,30	-0,10	0,08	0,08	0,14	0,17	-0,10	0,08	-0,62	0,36	0,20	-0,55	-0,23	-0,23	0,24	0,36	0,36	0,17	0,30	0,04	-0,22	0,08	0,34	0,04	-0,13	-0,17	-0,41	0,14	0,38	0,04			
312	0,10	0,10	0,10	0,45	0,10	0,19	-0,35	0,23	-0,35	-0,35	-0,17	-0,31	-0,16	-0,35	0,30	-0,59	0,70	0,45	-0,03	-0,03	-0,12	-0,59	-0,59	0,19	-0,35	0,04	0,02	-0,35	-0,18	0,04	0,12	-0,03	0,37	0,33	0,03	0,04			
321	-0,29	-0,29	-0,29	-0,06	-0,29	0,10	0,10	0,03	0,10	0,10	0,18	0,07	-0,11	0,10	-0,66	0,18	0,19	-0,62	-0,24	-0,24	-0,09	0,18	0,18	0,10	0,10	0,29	-0,34	0,10	0,37	0,29	-0,06	-0,12	-0,40	0,38	0,38	0,29			
322	-0,33	-0,33	-0,33	-0,04	-0,33	0,16	0,10	-0,09	0,10	0,10	0,12	0,11	-0,06	0,10	-0,66	0,28	0,16	-0,62	-0,26	-0,26	0,09	0,28	0,28	0,16	0,10	0,13	-0,37	0,10	0,40	0,13	-0,01	-0,06	-0,44	0,27	0,46	0,13			
323	0,46	0,46	0,46	0,09	0,46	-0,35	-0,39	0,27	-0,28	-0,28	-0,26	-0,22	0,20	-0,28	0,66	-0,42	-0,16	0,39	0,23	0,23	-0,24	-0,42	-0,42	-0,35	-0,39	0,11	0,20	-0,28	-0,40	0,11	-0,04	0,01	0						

Tabelle A.XXII: Euklidische Distanzen der Aminosäurepositionen aus Matrix 3

	7	23	24	42	54	57	83	93	104	162	173	190	201	205	219	221	230	237	246	247	251	256	259	265	267	269	270	273	281	282	287	288	293	295	297			
23	0,0																																					
24	22,7	22,7																																				
42	45,3	45,3	28,4																																			
54	35,1	35,1	57,9	77,5																																		
57	91,0	91,0	113,7	131,2	56,2																																	
83	85,1	85,1	101,8	120,5	67,0	72,6																																
93	51,9	51,9	62,9	80,2	52,7	89,5	70,0																															
104	88,0	88,0	110,3	127,1	54,3	21,9	75,4	88,6																														
162	86,4	86,4	108,2	124,3	54,1	29,5	77,5	88,1	8,3																													
173	91,7	91,7	114,4	132,4	56,6	11,3	73,0	90,7	17,0	25,2																												
190	31,7	31,7	28,2	36,1	58,0	110,3	98,1	72,4	105,6	103,0	110,5																											
201	30,2	30,2	20,8	33,5	60,7	114,4	102,6	70,8	109,1	106,1	114,7	18,6																										
205	97,2	97,2	111,7	120,5	82,6	87,9	111,0	106,0	68,1	59,7	84,7	103,8	103,3																									
219	25,6	25,6	43,2	60,4	28,6	77,5	78,4	49,4	75,2	73,6	79,7	46,4	47,1	88,7																								
221	104,0	104,0	118,4	127,1	89,0	91,1	103,4	116,2	89,8	89,9	91,0	103,2	112,5	110,7	95,9																							
230	120,2	120,2	127,2	130,5	117,5	133,2	134,3	132,4	130,7	129,7	133,3	111,8	120,3	137,5	115,8	49,7																						
237	175,7	175,7	176,3	187,2	180,6	201,3	211,4	196,1	202,8	203,2	202,1	176,5	177,4	215,7	178,4	211,5	222,5																					
246	45,7	45,7	68,0	85,0	14,4	47,2	66,4	58,1	46,4	46,7	48,6	64,4	69,0	80,0	35,2	80,3	111,4	185,1																				
247	29,0	29,0	24,9	29,2	56,8	108,8	101,1	64,6	106,1	103,9	110,5	22,9	25,0	107,3	40,8	104,4	111,1	181,7	62,1																			
251	71,9	71,9	91,0	106,7	49,0	55,7	77,5	89,0	56,3	57,7	56,5	85,2	91,2	92,2	61,0	61,6	97,3	196,3	43,7	85,0																		
256	169,5	169,5	174,1	173,5	168,2	180,2	178,7	185,4	177,9	176,8	180,2	156,5	165,5	180,4	166,4	89,3	58,7	253,5	162,1	158,3	143,0																	
259	90,7	90,7	109,7	122,7	66,1	77,8	82,5	98,9	57,6	59,1	57,4	97,9	106,0	93,5	79,8	33,7	79,1	203,7	55,7	98,5	40,7	123,0																
265	101,4	101,4	113,4	120,0	91,3	93,1	112,7	107,2	100,7	100,7	102,6	111,6	110,9	119,5	88,8	120,8	144,5	207,1	85,0	100,7	99,8	185,9	105,9															
267	92,6	92,6	105,9	123,6	80,8	91,9	19,5	75,7	94,3	90,8	92,5	102,5	106,6	123,6	88,2	114,2	140,5	217,7	81,7	106,2	92,3	182,8	97,5	122,7														
269	158,3	158,3	157,8	165,7	165,3	190,1	189,7	139,0	188,4	187,4	190,5	168,7	164,4	191,6	157,7	207,5	200,0	245,5	168,0	157,9	188,1	255,2	197,2	195,0	193,7													
270	54,5	54,5	66,8	82,9	51,9	85,7	65,9	48,4	84,6	84,1	86,8	71,1	70,0	102,8	47,0	102,7	128,6	195,4	57,2	69,1	82,1	174,7	89,6	105,8	72,1	176,1												
273	135,0	135,0	141,6	143,0	132,1	145,8	159,2	146,5	126,2	117,8	142,8	133,4	130,1	58,1	130,7	152,2	166,7	241,7	131,7	138,1	142,7	201,6	143,9	158,6	167,5	212,4	144,3											
281	117,3	117,3	107,4	85,8	138,7	181,7	172,6	144,4	177,2	174,6	182,2	98,8	107,4	167,0	127,5	151,4	142,1	227,3	140,5	97,0	151,9	167,7	158,4	172,5	174,3	208,3	146,0	179,4										
282	126,4	126,4	129,7	140,7	129,2	151,8	154,9	106,3	150,3	149,6	152,1	139,6	136,2	158,4	124,0	174,1	172,6	224,9	131,3	129,1	152,0	230,0	161,0	163,4	160,9	38,4	143,2	185,9	189,0									
287	90,9	90,9	76,1	61,7	118,9	167,2	159,5	125,0	161,3	157,3	168,4	69,3	66,2	140,3	105,3	153,8	148,8	181,4	124,1	72,2	145,0	184,0	154,8	139,7	161,5	183,1	128,3	146,9	101,8	164,0								
288	75,7	75,7	65,0	56,7	100,5	146,8	142,3	110,5	141,0	137,0	148,2	56,6	52,8	122,3	86,8	135,1	133,7	174,0	105,0	59,0	125,5	172,0	134,6	121,3	145,8	176,7	113,3	134,5	109,5	154,1	27,4							
293	149,7	149,7	131,3	122,7	180,1	230,9	214,1	161,8	227,2	232,6	150,2	140,5	215,5	225,2	213,8	237,7	189,6	141,2	203,3	249,8	224,6	209,1	212,4	182,9	167,0	221,8	139,9	165,9	145,7	145,8	123,7	149,1						
295	69,2	69,2	77,2	78,4	70,8	102,7	108,2	89,0	100,6	99,8	103,1	74,3	82,9	113,5	73,4	124,4	144,9	197,9	73,3	74,9	95,2	189,4	109,2	120,6	116,9	175,2	98,2	150,2	118,8	144,6	111,3	103,7	184,0					
297	76,3	76,3	97,8	112,7	45,6	30,1	71,8	81,0	36,8	39,7	35,8	93,0	98,0	85,0	63,4	87,5	125,9	195,4	36,0	90,9	52,2	172,9	58,3	77,6	89,2	183,2	80,0	139,9	165,9	145,7	145,8	123,5	217,3	83,2				
302	84,1	84,1	106,4	125,1	50,7	24,2	72,7	77,2	28,6	34,0	23,2	105,2	108,6	87,3	71,3	94,6	130,8	199,2	43,5	102,6	58,2	182,4	62,7	100,6	91,3	168,0	82,7	144,1	177,6	129,6	162,4	142,5	220,5	97,8	37,4			
303	84,1	84,1	106,4	122,8	51,3	28,8	73,1	85,8	30,9	35,8	26,4	101,9	107,5	87,5	73,0	95,5	135,4	198,8	44,6	102,7	59,1	182,6	64,0	104,8	91,5	187,3	82,3	143,9	169,6	149,4	160,1	143,0	224,3	87,0	41,2			
304	72,6	72,6	66,0	58,7	93,3	135,9	134,3	102,1	134,8	132,4	139,5	62,2	63,6	131,1	77,1	142,3	150,2	170,3	96,7	58,0	122,5	190,7	135,3	105,4	138,7	181,1	102,8	153,9	103,2	156,2	66,7	65,1	149,6	95,7	117,9			
305	78,9	78,9	93,9	117,9	51,3	45,9	68,4	78,4	33,6	29,1	46,7	99,8	103,0	95,0	67,5	93,7	150,3	198,6	49,2	98,8	34,4	177,1	66,4	102,9	94,9	187,2	87,3	148,4	170,1	150,3	159,0	140,1	209,1	96,5	49,5			
308	82,9	82,9	105,0	123,8	50,1	27,3	73,1	75,3	31,0	35,8	26,5	104,2	107,4	87,5	70,2	95,3	130,4	198,8	43,2	101,3	58,9	182,6	63,9	100,7	91,5	164,6	82,2	143,8	176,6	126,2	161,1	141,3	218,5	97,1	38,6			
310	61,5	61,5	82,5	98,1	34,1	46,1	59,7	68,1	46,2	47,7	45,6	78,1	83,4	85,5	56,6	76,7	77,4	118,8	197,1	28,3	77,3	46,7	160,3	51,9	95,1	75,5	178,3	66,3	138,1	146,6	141,5	137,6	118,4	204,2	78,3	39,9		
312	127,2	127,2	120,5	113,1	144,2	178,5	178,6	150,5	180,2	178,7	183,3	119,3	119,6	179,4	130,0	187,3	193,2	174,1	147,0	116,1	169,4	227,8	181,3	135,9	181,8	211,2	152,0	198,1	141,8	191,5	101,6	110,9	173,1	129,9	161,9			
321	113,6	113,6	100,3	84,7	139,1	185,8	159,1	131,6	178,1	174,0	185,5	98,5	96,8	154,4	128,6	170,8	164,0	238,2	144,8	100,3	162,8	196,7	172,4	171,0	155,7	192,0	140,4	156,3	102,0	175,5	85,6	90,7	150,5	121,1	166,0			
322	98,8	98,8	83,6	64,8	126,7	175,7	152,7	121,6	169,3	165,4	176,4	83,7	81,6	148,9	112,9	161,2	154,8	222,9	132,7	82,8	150,2	189,3	162,5	156,1	150,6	184,7	128,4	154,8	86,6	167,1	70,9	76,1	132,9	112,3	155,6			

Erklärung

Die hier vorgelegte Dissertation habe ich eigenständig und ohne unerlaubte Hilfe angefertigt. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.