

Auswirkungen von räumlichem Populationswachstum auf die genetische Vielfalt

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

“Doctor rerum naturalium” (Dr. rer. nat.)

der Georg-August-Universität Göttingen

Im Grundprogramm Biologie
der Georg-August University School of Science (GAUSS)

vorgelegt von

Sven Boekhoff

aus Westerstede

Göttingen, 2014

Betreuungsausschuss:

Prof. Dr. Dirk Semmann; CRC Nachwuchsgruppe „Evolution von Kooperation und Prosozialem Verhalten“, Courant Forschungszentrum „Evolution des Sozialverhaltens“.

Prof. Dr. Kerstin Wiegand; Abteilung „Ökosystemmodellierung“, Büsgen-Institut, Georg-August-Universität Göttingen.

Dr. Oskar Hallatschek; Max-Planck-Forschungsgruppe „Biologische Physik und Evolutionäre Dynamik“, Max-Planck-Institut für Dynamik und Selbstorganisation.

Mitglieder der Prüfungskommission:

Erstreferentin: Prof. Dr. Kerstin Wiegand

Zweitreferent: Prof. Dr. Dirk Semmann

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Holger Kreft; Free Floater Junior Research Group „Biodiversity, Macroecology and Conservation Biogeography“, Fakultät für Forstwissenschaften und Waldökologie, Georg-August-Universität Göttingen.

Prof. Dr. Ulrich Brose; „Systemische Naturschutzbiologie“, J.F. Blumenbach Institut für Zoologie und Anthropologie, Georg-August-Universität Göttingen.

Prof. Dr. Julia Ostner; CRC Nachwuchsgruppe „Soziale Evolution bei Primaten“, Courant Forschungszentrum „Evolution des Sozialverhaltens“.

Dr. Oliver Schülke; Courant Forschungszentrum „Evolution des Sozialverhaltens“.

Tag der mündlichen Prüfung: 01.08.2014

... der Erkenntnis

Inhaltsverzeichnis

Zusammenfassung	vii
1. Einleitung	1
1.1. Beispiele für Range-Expansions	2
1.2. Ausbreitungsgeschichte des Menschen	3
1.3. Einfluss von Range-Expansions auf die genetische Zusammensetzung von Populationen	7
1.4. Zwei Typen genetischer Variation	10
1.5. Ökologische und geographische Faktoren	11
1.6. Zielsetzung und Struktur dieser Arbeit	13
2. Methoden	15
2.1. Simulationsmodell	17
2.1.1. Geographie/Gitter	17
2.1.2. Eigenschaften der Deme	19
2.1.3. Arten und Verteilung der genetischen Variation	28
2.1.4. Demographische Szenarien	29
2.2. Umsetzung des Modells	31
2.2.1. Laufzeiten und Wiederholungen	31
2.2.2. Messgrößen	31
3. Ergebnisse	33
3.1. Einfluss demographischer Szenarien auf Allelfrequenz-Spektren	34
3.2. Einfluss von Mutationsrate und Anfangsfrequenz auf Allelfrequenz-Spektren	35
3.3. Beitrag von Standing-genetic-Variation und Ongoing-Mutations zu den Allelfrequenz-Spektren	38
3.4. Einfluss der geographischen Struktur auf Allelfrequenz-Spektren	42
3.5. Potenzgesetz in Simulationen auf der Großen Welt	45
4. Diskussion	53
4.1. Range-Expansions generieren Allelfrequenz-Spektren, die einem Potenzgesetz folgen	54
4.2. Mutationen aus der Ursprungspopulation dominieren Allelfrequenz-Spektren	58
4.3. Anwendung auf echte genetische Daten	60
4.4. Ausblick	62

5. Literaturverzeichnis	65
A. Anhang (Methoden)	75
A.1. Details zum Modell	75
A.2. Implementierungen	77
A.2.1. Implementierung von Selektion	77
A.2.2. Implementierung des Populationswachstums	82
A.2.3. Implementierung des Allee Effektes	83
A.2.4. Implementierung der Zufallsverteilungen	85
A.3. Auswertungen	90
A.3.1. Allelfrequenz-Spektren	90
A.3.2. Echte genetische Daten	96
B. Anhang (Ergebnisse)	99
B.1. Simulationsergebnisse	99
B.1.1. Vergleich demographischer Szenarien	99
B.1.2. Vergleich von Standing-genetic-Variation und Ongoing-Mutations	99
B.1.3. Einfluss der Geographie auf die Allelfrequenz-Spektren	105
B.1.4. Potenzgesetz in Simulationen auf der Großen Welt	105
B.2. Theoretische Erwartung für den Exponenten des Potenzgesetzes	113
B.3. Allelfrequenz-Spektren aus echten genetischen Daten	116
C. Anhang (Bildquellen)	119
Lebenslauf	121
Danksagung	123

Zusammenfassung

Wächst eine Population und breitet sich dabei geographisch aus, so spricht man von räumlichem Populationswachstum, bzw. einer *Range-Expansion*. Viele Arten haben im Verlaufe ihrer evolutionären Geschichte ihr Verbreitungsgebiet ausgeweitet. Gründe hierfür können beispielsweise ein geändertes Klima oder die Verschleppung der Art in einen neuen Lebensraum sein.

Während einer Range-Expansion können durch *Gene-Surfing* räumliche Verteilungen von neutralen genetischen Varianten entstehen, die den Folgen von selektiven Prozessen ähnlich sind. Für eine korrekte Interpretation der genetischen Daten ist daher die Kenntnis über quantitative Auswirkungen von Range-Expansions auf die genetische Vielfalt unumgänglich.

In dieser Arbeit charakterisiere ich die Konsequenzen von Range-Expansions für Allelfrequenz-Spektren. Dazu generiere ich in Computersimulationen genetische Daten für unterschiedliche demographische Szenarien sowie diverse ökologische und geographische Bedingungen.

Ich zeige, dass Range-Expansions innerhalb kurzer Zeit zu Allelfrequenz-Spektren führen können, die sich durch ein Potenzgesetz mit einem spezifischen Exponenten beschreiben lassen. Dieser Exponent liegt zwischen den erwarteten Exponenten für stabile und exponentiell wachsende, durchmischte Populationen. Mutationen, die während einer Range-Expansion aufgetreten sind, tragen meinen Ergebnissen zufolge weniger zu heutigen Allelfrequenz-Spektren bei, als Mutationen, die bereits in der Ursprungspopulation vorhanden waren. Allerdings eignen sich neue Mutationen besser, um Range-Expansions in genetischen Daten aufzuspüren, da sie weniger von geographischen Strukturen beeinflusst werden.

Meine Resultate werden dazu beitragen, Spuren von Range-Expansions in genetischen Daten zu entdecken und Rückschlüsse auf die evolutionäre Vergangenheit von Populationen zu ziehen.

1. Einleitung

Das Leben auf der Erde entwickelt sich seit Milliarden von Jahren (Wacey et al., 2011) und hat im Laufe der Zeit eine Vielzahl komplexer Lebensformen hervorgebracht (Mora et al., 2011). Populationen von Tieren, Pflanzen und Mikroorganismen haben selbst extreme Lebensräume auf der Erde besiedelt. Wächst eine solche Population und kolonisiert dabei angrenzende Gebiete, so spricht man von *räumlichem Populationswachstum* bzw. einer Range-Expansion (Excoffier et al., 2009). Range-Expansions hat es in vielen, wenn nicht sogar in jeder Pflanzen- und Tierart gegeben (Hewitt, 2000; Davis et al., 2011; Chen et al., 2011). In ABSCHN. 1.1 führe ich zunächst einige Beispiele und Ursachen für Range-Expansions aus der Tier- und Pflanzenwelt auf, um dann in ABSCHN. 1.2 im Detail auf die Ausbreitungsgeschichte des Menschen einzugehen.

Die Migrationsgeschichte vieler Arten wurde bisher vor allem mit Hilfe fossiler Funde untersucht. In Studien konnte gezeigt werden, dass Range-Expansions einen wesentlichen Einfluss auf die genetische Zusammensetzung und damit auf das Schicksal einer Population haben können (Olivieri, 2009; Excoffier and Ray, 2008). Diese Spuren von Range-Expansions in genetischen Daten werden vermehrt verwendet, um das bisherige Wissen über die Geschichte von Populationen zu ergänzen (Linz et al., 2007; Nielsen et al., 2007; Reich et al., 2010). In ABSCHN. 1.3 erläutere ich die Eigenschaften und die populationsgenetischen Ursachen von genetischen Veränderungen während einer Range-Expansion. Darüber hinaus gehe ich auf Probleme bei der Interpretation genetischer Daten ein, die kürzlich anhand von Computersimulationen entdeckt wurden (Nielsen et al., 2007; Excoffier et al., 2009; Currat et al., 2006; Hofer et al., 2009).

Ziel dieser Arbeit ist es, den quantitativen Einfluss von Range-Expansions auf die genetische Zusammensetzung von Modellpopulationen zu analysieren. Dabei ist, wie ich in den folgenden Kapiteln zeigen werde, die Unterscheidung von bestehender Diversität (*standing genetic variation*) und neu auftretenden Mutationen (*ongoing mutation*) von besonderer Bedeutung. Unter *Standing-genetic-Variation* verstehe ich dabei die genetische Variation, die innerhalb der Ursprungspopulation vorhanden ist, also bevor die

1. Einleitung

Range-Expansion beginnt. Mutationen, die erst während der Range-Expansion entstehen, bezeichne ich als *Ongoing-Mutations* (siehe ABSCHN. 1.4).

In Computersimulationen habe ich verschiedene demographische Szenarien sowohl mit Ongoing-Mutations als auch mit Standing-genetic-Variation in vereinfachten (idealisierten) Populationen nachgestellt. Um in meinen Simulationen realistische Bedingungen zu modellieren, habe ich Simulationen auf einer Europakarte durchgeführt. Darüber hinaus habe ich den Einfluss der sogenannten Allee-Effektes (ABSCHN. 1.5) in meine Simulationen einbezogen.

1.1. Beispiele für Range-Expansions

Wie die Beispiele in ABB. 1.1 zeigen haben Range-Expansions in einem breiten Spektrum biologischer *Taxa* stattgefunden, bzw. finden immer noch statt. Anfängen von Bakterien ABB. 1.1(F) über wirbellose Tiere ABB. 1.1(B,E) und Pflanzen ABB. 1.1(D) bis hin zu Wirbeltieren ABB. 1.1(A,C). In der menschlichen Wahrnehmung erscheinen die Verbreitungsräume vieler Arten jedoch weitgehend stabil. Das hängt unter anderem mit den Zeit- und Längenskalen zusammen, in denen Range-Expansions passieren. So wachsen etwa Biofilme zwar in der Größenordnung von Stunden oder Tagen, entziehen sich allerdings aufgrund ihrer geringen Größe der alltäglichen Beobachtung (Hallatschek et al., 2007). Im Gegensatz dazu lassen sich die Besiedlung durch invasive Arten im Laufe von Jahrzehnten (Thompson et al., 1987; Phillips and Shine, 2005; Hebert et al., 1989) oder die Ausbreitung des Menschen in der Welt innerhalb von Jahrtausenden (Oppenheimer, 2012; Pinhasi et al., 2012; Stoneking and Krause, 2011) aufgrund ihrer Dauer nicht direkt beobachten.

Die *Ursachen* für Range-Expansions können dabei sehr vielfältig sein. Neben ökologischen Faktoren wie dem Nahrungsangebot oder Konkurrenzsituationen können auch abiotische Faktoren wie z.B. ein sich veränderndes Klima Auslöser für Range-Expansions sein (Kubisch et al., 2013; Chen et al., 2011). So hat z.B. der gemeine Grashüpfer (*Chorthippus parallelus*, ABB. 1.1(B)) in Folge von Klimaoszillationen sein Verbreitungsgebiet in Europa mehrfach geändert (Hewitt, 1996, 2004). Nach Gebietsverkleinerungen (*range contraction*) während vergangener Eiszeiten besiedelte er nördliche Regionen vermutlich wiederholte Male aus Refugien in Südeuropa (Hewitt, 2000, 1996). Auch bei der Besiedlung Europas durch den Menschen wird ein starker Einfluss der letzten Eiszeit angenommen (siehe dazu ABSCHN. 1.2).

Neben Klimaveränderungen ist die Verschleppung von Arten (vor allem durch den Menschen) eine weitere Ursache für Range-Expansions. Die Ökologie eingeführter invasiver Arten hat sich zu einem vielbeachteten Forschungszweig entwickelt, insbesondere wegen ihrer negativen Einflüsse auf die biologische Vielfalt und die Wirtschaft (Davis et al., 2011). So verursacht etwa die Zebrauschel (*Dreissena polymorpha*; ABB. 1.1(E)), eine von rund 50'000 in die USA eingeführten Arten, allein dort einen wirtschaftlichen Schaden von jährlich etwa einer Milliarde US-Dollar (Pimentel et al., 2005). Eine Bedrohung einheimischer Pflanzenarten stellt der ebenfalls nach Nordamerika eingeschleppte gewöhnliche Blutweiderich (*Lythrum salicaria*; ABB. 1.1(D)) dar (Thompson et al., 1987). Durch sein dominantes Wachstum verringert dieser Neophyt die Biodiversität in Feuchtgebieten.

Abgesehen von äußeren Einflüssen, wie Klima oder Verschleppung, können auch Anpassungen (Adaptationen) zur erfolgreichen Ausbreitung einer Spezies führen. In ihrer Studie aus dem Jahr 2010 konnten Bocxlaer et al. zeigen, dass die Ausbreitung der Kröten (*Bufo*) vor etwa 30 Millionen Jahren durch die Entstehung eines Range-Expansion Phänotyps ermöglicht wurde. Auch Anpassungen an klimatische Bedingungen außerhalb des ursprünglichen Verbreitungsraumes konnten in expandierenden Arten nachgewiesen werden. So ermöglichen Änderungen in der Temperaturtoleranz der Wespenspinne (*Argiope bruennichi*; Krehenwinkel and Tautz (2013)) und eine zeitlich versetzte Blütezeit in Gemeinem Blutweiderich (Colautti and Barrett, 2013) diesen Arten die Besiedlung kälterer Regionen.

Bei gegenwärtig invasiven Arten können die Auswirkungen demographischer Änderungen auf genetische Daten unmittelbar beobachtet werden. Daher eignen sich solche Arten als Modellsysteme zur Untersuchung von Range-Expansions (Allendorf and Lundquist, 2003). Ziel ist es, aus empirischen Daten Informationen über die demographische Geschichte von Populationen (z.B. die Besiedlung Europas durch den Menschen) zu gewinnen.

1.2. Ausbreitungsgeschichte des Menschen

Es herrscht ein breiter Konsens darüber, dass die Besiedlung der Welt durch den anatomisch modernen Menschen vor etwa 70 Tausend Jahren in Afrika begann (Out-Of-Africa Hypothese; Stoneking and Krause (2011); Oppenheimer (2012)). Jüngste Fortschritte bei den Sequenzieretechniken und in der Paläogenetik ergänzen unser bisheriges Bild

1. Einleitung

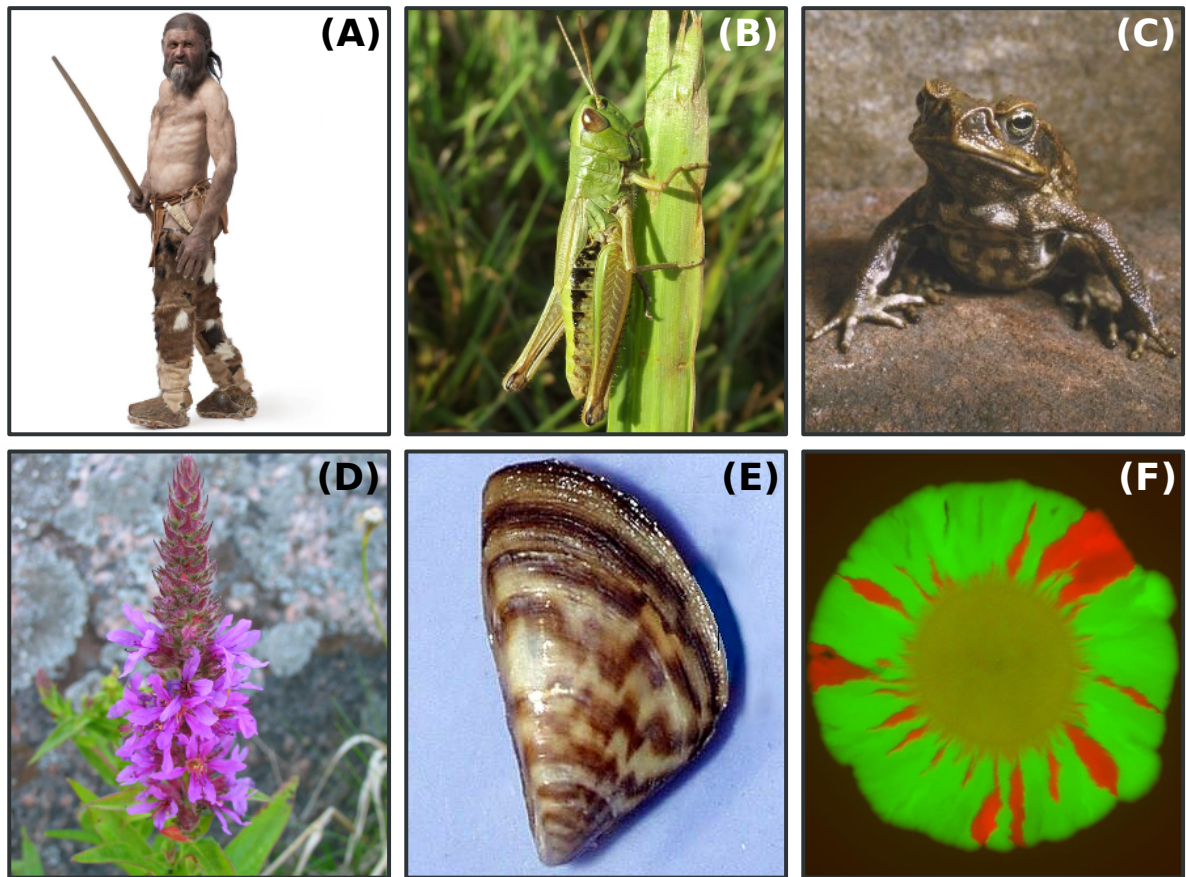


Abbildung 1.1.: **Beispiele für Range-Expansions.** Range-Expansions gab es vermutlich in der evolutionären Geschichte jeder Art. (A) Ausgehend von Afrika hat der anatomisch moderne Mensch die Welt vor etwa 70'000 Jahren besiedelt ([Oppenheimer, 2012](#)). Das Bild zeigt die Rekonstruktion des 'Mannes vom Hauslabjoch' (auch 'Ötzi' genannt), der vor etwa 5'000 Jahren gelebt hat. (B) In Folge von Klimaveränderungen, hat der Gemeine Grashüpfer (*Chorthippus parallelus*) den Europäischen Kontinent mehrfach besiedelt ([Hewitt, 2000, 1996](#)). (C) Die Aga-Kröte (*Bufo marinus*) ist eine von mehr als 50 eingeschleppten Arten in Australien. Sie wurde in den 30er Jahren des vergangenen Jahrhunderts zur Bekämpfung landwirtschaftlicher Schädlinge eingeführt, und breitet sich seither stetig aus ([Easteal, 1981](#)). (D) Der Gewöhnliche Blutweiderich (*Lythrum salicaria*) wurde bereits zu Beginn des 19. Jahrhunderts auf verschiedene Weise nach Nordamerika eingeführt ([Thompson et al., 1987](#)). Seit etwa 1930 breitet sich dieser Neophyt aggressiv aus und verdrängt in Feuchtgebieten einheimische Pflanzen. (E) Die aus Südrussland stammende Zebrauschel (*Dreissena polymorpha*) gelangte vermutlich über Ballastwasser von Schiffen Mitte der 1980er Jahre nach Nordamerika ([Hebert et al., 1989](#)). Dort richtet sie durch die Besiedlung von Oberflächen große wirtschaftliche Schäden an. (F) Wachsende Bakterienkolonie, die aus einem Gemisch (jeweils 50%) zweier *Escherichia coli*-Stämme hervorging. Jeder Stamm ist mit einem neutralen genetischen Marker versehen, der im Fluoreszenzmikroskop sichtbar gemacht werden kann. Während des Wachstums bilden sich Regionen aus, in denen jeweils nur einer der Stämme vorkommt ([Hallatschek et al., 2007](#)). Weitere Informationen und Referenzen zu den einzelnen Arten befinden sich im Haupt-Text. Für Bildquellen siehe KAP. C.

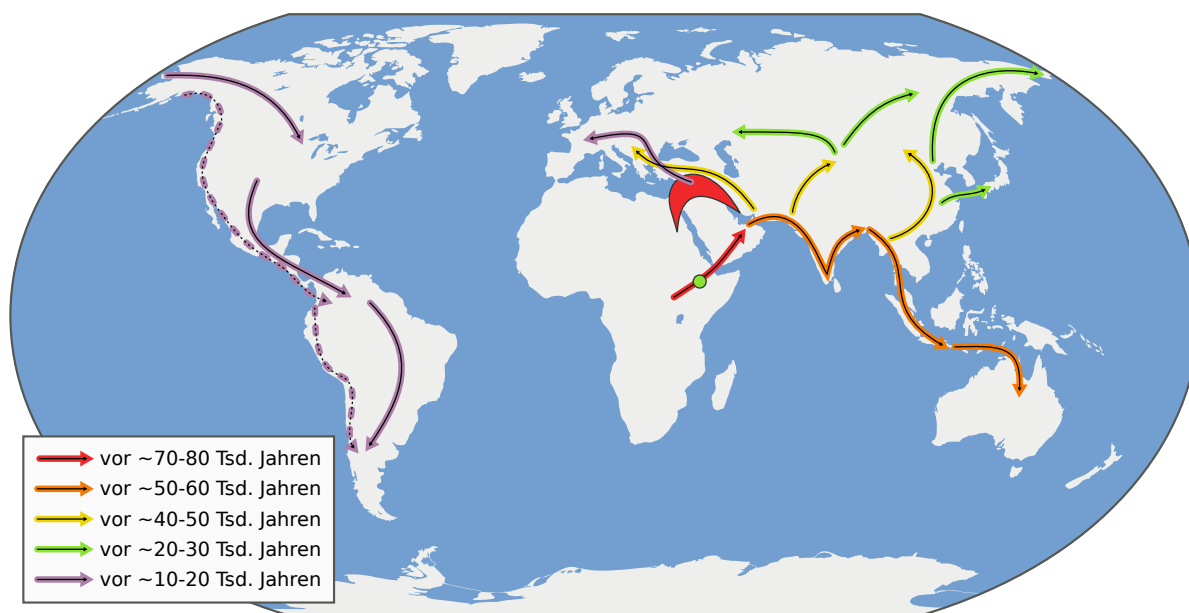


Abbildung 1.2.: **Vermutliche Verbreitungswege des anatomisch modernen Menschen.** Ausgehend von einer Region um Addis Abeba (grüner Kreis) begann vor etwa 70 Tausend Jahren die Ausbreitung des anatomisch modernen Menschen in nahezu alle Teile der Welt (Oppenheimer, 2012). Der genaue räumliche und zeitliche Verlauf der Wanderungsbewegungen ist noch unklar. Fortschritte in den Sequenzieretechniken und neue Erkenntnisse aus paläogenetischen Studien deuten aber darauf hin, dass es nur ein einziges Out-of-Africa Ereignis des modernen Menschen gegeben hat (Green et al., 2010). Der europäische Kontinent wurde durch den modernen Menschen vermutlich während des Paläolithikum (vor etwa 40'000 Jahren) und ein zweites Mal nach der letzten Eiszeit vor etwa 10'000 Jahren besiedelt (*neolithische Expansion*). Es wird angenommen, dass der Ausgangspunkt beider Expansionen in der Region des fruchtbaren Halbmondes (rote Fläche) lag. Bildvorlage: KAP. C.

menschlicher Migrationswege bei der Besiedlung der Welt (ABB. 1.2). Dabei wird unter anderem der Beitrag von solchen Genen zum heutigen Genpool betrachtet, die durch Introgression aus Populationen archaischer Menschen eingebracht wurden. So fanden Green et al. (2010), dass alle Populationen außerhalb Afrikas einen ähnlichen Anteil (1 – 4%) an Neandertaler-Genen aufweisen. Sie schlossen daraus, dass diese Gene bereits vor dem Beginn der Ausbreitung in die Population eingebracht wurden und es vermutlich nur ein einziges Out-of-Africa Event gegeben hat.

Die Verbreitung von Genen des Denisova-Menschen in heutigen Populationen deutet darauf hin, dass der asiatische und ozeanische Raum mehrfach besiedelt wurde (Reich et al., 2010, 2011). Durch die Sequenzierung von DNA einer seit mehr als 12'000 Jahre erhaltenen Leiche konnten Rasmussen et al. (2014) darauf schließen, dass die Vorfahren

1. Einleitung

der amerikanischen Ureinwohner vor mehr als 13'000 Jahren nach Amerika einwanderten. Zudem konnten [Rasmussen et al. \(2010\)](#) zeigen, dass Teile Nordamerikas (ausgehend von einer sibirischen Population) vor etwa 5'000 Jahren ein zweites Mal besiedelt worden waren. Die genauen Migrationswege und deren zeitliche Einordnung wird jedoch weiterhin debattiert. So führten jüngste Messungen der Mutationsrate im menschlichen Genom zu der Erkenntnis, dass die Migrationsbewegungen der Menschen vor deutlich längerer Zeit stattgefunden haben, als bisher angenommen wurde ([Conrad et al., 2011](#); [Scally and Durbin, 2012](#)).

Die Besiedlung Europas wird in der Regel in drei entscheidende Phasen unterteilt ([Pinhasi et al., 2012](#)). Ausgehend von einer Region am fruchtbaren Halbmond (siehe [ABB. 1.2](#)) fand vor etwa 40'000 Jahren eine Ausbreitung paläolithischer Jäger- und Sammlerpopulationen nach Europa statt (*Paläolithische Expansion*). Im Anschluss gab es Bereiche Europas, die durch die klimatischen Bedingungen während der letzten *Eiszeit* (vor etwa 27'000 – 16'000 Jahren) nicht bewohnbar waren. Vor etwa 10'000 Jahren fand vermutlich ein erneutes Wachstum der menschlichen Population statt, ausgelöst durch die Entwicklung der Landwirtschaft in der Region zwischen Euphrat und Tigris in der heutigen Türkei und Syrien ([Lev-Yadun et al., 2000](#)). In Folge dieser Entwicklungen wurden die paläolithischen Jäger und Sammler durch neolithische Bauern ersetzt (*Neolithische Revolution* bzw. *neolithic transition*).

Es werden vor allem zwei Modelle diskutiert, die den Ablauf der Neolithischen Revolution erklären könnten ([Pinhasi et al., 2012](#)). Das *Diffusionismus-Modell* (*cultural diffusion model*; CD) geht davon aus, dass sich das Wissen und die Nutzpflanzen ohne eine Wanderung von Populationen ausgebreitet haben. Anhänger des Modells stützen ihre Aussage vor allem auf den großen Anteil (vor allem mitochondrialer) genetischer Marker aus Populationen, von denen vermutlich die paläolithische Expansion nach Europa ausging ([Haak et al., 2005](#); [Belle et al., 2006](#)). Das *Landnahme-Modell* (*demic diffusion model*; DD) geht davon aus, dass sich die Landwirtschaft durch die Ausbreitung der Bauernpopulationen in Europa verbreitet hat. Für dieses Modell sprechen fossile Funde sowie Ergebnisse aus der Analyse von Y-Chromosom-Markern (siehe Referenzen in [Pinhasi et al. \(2012\)](#)). Darüber hinaus konnten [Currat and Excoffier \(2005\)](#) mit einer Simulationsstudie zeigen, dass ein großer Anteil an paläolithischen Genen auch während Wanderungsbewegungen durch Introgression in den Genpool aufgenommen worden sein könnte. Obwohl es sehr wahrscheinlich genetischen Austausch zwischen paläolithischen und neolithischen Populationen gegeben hat, ist der tatsächliche Anteil paläolithischer Gene am heutigen europäischen Genpool unklar ([Pinhasi et al., 2012](#); [Belle et al., 2006](#);

1.3. Einfluss von Range-Expansions auf die genetische Zusammensetzung von Populationen

[Semino et al., 2000](#)).

Um mit Hilfe empirischer Daten die komplexe Migrationsgeschichte von Populationen verstehen zu können, müssen zunächst grundlegende Einflüsse der demographischen Geschichte auf die genetische Vielfalt näher untersucht werden. Ziel dieser Arbeit ist es, die bisher wenig erforschten quantitativen Einflüsse von Range-Expansions ([Excoffier et al., 2009](#)) genauer zu analysieren.

1.3. Einfluss von Range-Expansions auf die genetische Zusammensetzung von Populationen

Während einer Range-Expansion geht die genetische Vielfalt mit zunehmender Entfernung von der Ausgangspopulation verloren ([Austerlitz et al., 1997](#)). Ursache dafür sind aufeinanderfolgende Gründereffekte (*serial founder effects*) (siehe [ABB. 1.3](#) und [Slatkin and Excoffier \(2012\)](#)). Diese Spuren von Range-Expansions konnten auch in echten genetischen Daten nachgewiesen werden. So fanden [Prugnolle et al. \(2005\)](#) in einer Studie mit 377 genetischen Markern, dass die genetische Vielfalt in menschlichen Populationen mit zunehmender Entfernung von Afrika abnimmt. Diese stetige Abnahme der Vielfalt konnte auch für phänotypische Merkmale an Schädeln ([Manica et al., 2007](#)), Laute in gesprochenen Sprachen (Phoneme) ([Atkinson, 2011](#)) und die genetische Vielfalt des im menschlichen Magen vorkommenden Bakteriums *Helicobacter pylori* ([Linz et al., 2007](#)) nachgewiesen werden.

In einer wegweisenden Studie aus dem Jahre 2004 untersuchten [Edmonds et al.](#) die Auswirkungen von Range-Expansions auf neutrale Mutationen, die während der Range-Expansion an der Front passieren (siehe dazu auch [ABB. 1.4\(A\)](#)). Dabei fanden sie heraus, dass in mehr als 20% ihrer Simulationen die neuen Mutationen nicht am Ort ihrer Entstehung, sondern auf halben Wege zwischen Entstehungsort und Populationsfront am häufigsten zu finden waren (*mutation travelling phenomenon*). Darüber hinaus erreichten einzelne Mutationen in seltenen Fällen während einer Range-Expansion extrem hohe Häufigkeiten in der Population. [Klopfstein et al. \(2006\)](#) konnten zeigen, dass dieses von ihnen als *Gene-Surfing* bezeichnete Phänomen stark davon abhängt, wie nahe die Mutation an der Front auftritt.

Die Ursache für Gene-Surfing ist, dass lediglich an der Front Gründereffekte und ein erhöhtes Populationswachstum für die Etablierung einer Mutation sorgen können,

1. Einleitung

welche die folgenden Neubesiedlungen dominiert (siehe dazu [ABB. 1.3](#)). In räumlichen Simulationen und Experimenten können auf diese Weise Bereiche (Sektoren) entstehen, in denen nur eine einzige genetische Variante vorkommt (siehe [ABB. 1.4\(A, B\)](#); [Edmonds et al. \(2004\)](#); [Hallatschek et al. \(2007\)](#)).

Eine solche lokale Anhäufung einer einzigen genetischen Variante wurde oftmals als Zeichen positiver Selektion interpretiert ([Nielsen et al., 2007](#)). Allerdings können derartige räumliche Verteilungen auch allein durch demographische Veränderungen und ohne die Notwendigkeit von Selektion entstehen ([Currat et al., 2006](#)). Mit einer Analyse verschiedener neutraler genetischer Marker konnte zudem gezeigt werden, dass vermutlich Gene-Surfing und nicht Selektion für die meisten der weltweit gefundenen Allelfrequenz-Unterschiede zwischen Populationen verantwortlich ist ([Hofer et al., 2009](#)). In einer Simulationsstudie aus dem Jahre 2007 konnten [Travis et al.](#) zeigen, dass auch nachteilige Mutationen auf der 'Welle der Ausbreitung' surfen und hohe Frequenzen erreichen können.

Neben den räumlichen Verteilungen einzelner genetischer Varianten werden auch *Allelfrequenz-Spektren* für die Charakterisierung populationsgenetischer Prozesse verwendet ([Nielsen et al., 2007](#)). Für ein Allelfrequenz-Spektrum ermittelt man die Frequenz der abgeleiteten Allele aller betrachteten Genorte in der Gesamtpopulation. Das Allelfrequenz-Spektrum gibt zu jeder Allelfrequenz die Anzahl der Genorte mit dieser Frequenz an. Ähnlich wie die räumlichen Verteilungen können auch Allelfrequenz-Spektren sowohl von demographischen Faktoren als auch von Selektionsprozessen beeinflusst werden ([Nielsen et al., 2007](#)). So erwartet man in einer Population, in der sich gerade eine sehr vorteilhafte Mutation ausbreitet (*selective sweep*), mehr Allele mit einer niedrigen Frequenz als in einer Population im Gleichgewicht. Grund hierfür ist der Verlust genetischer Diversität während des Sweeps sowie die kurze Zeit, in der neue Mutationen auftreten können ([Braverman et al., 1995](#); [Biswas and Akey, 2006](#); [Nielsen et al., 2007](#); [Excoffier et al., 2009](#)). Allerdings erwartet man diese Häufung von Allelen mit einer geringen Frequenz auch bei einer reinen demographischen Expansion (Wachstum einer Population ohne räumliche Ausbreitung; [Durrett \(2013\)](#)) und bei Range-Expansions (Populationswachstum mit räumlicher Ausbreitung; [Excoffier et al. \(2009\)](#)). [Coventry et al.](#) konnten 2010 erstmals diese Häufung seltener Allele in Allelfrequenz-Spektren von Menschen nachweisen, indem sie zwei Regionen des menschlichen Erbgutes in mehr als 13'000 Individuen sequenzierten. [Keinan and Clark \(2012\)](#) führten diese Häufung mit einer Simulationsstudie auf das exponentielle Wachstum in der menschlichen Population zurück, das in den letzten 400 Generationen stattgefunden hat.

1.3. Einfluss von Range-Expansions auf die genetische Zusammensetzung von Populationen

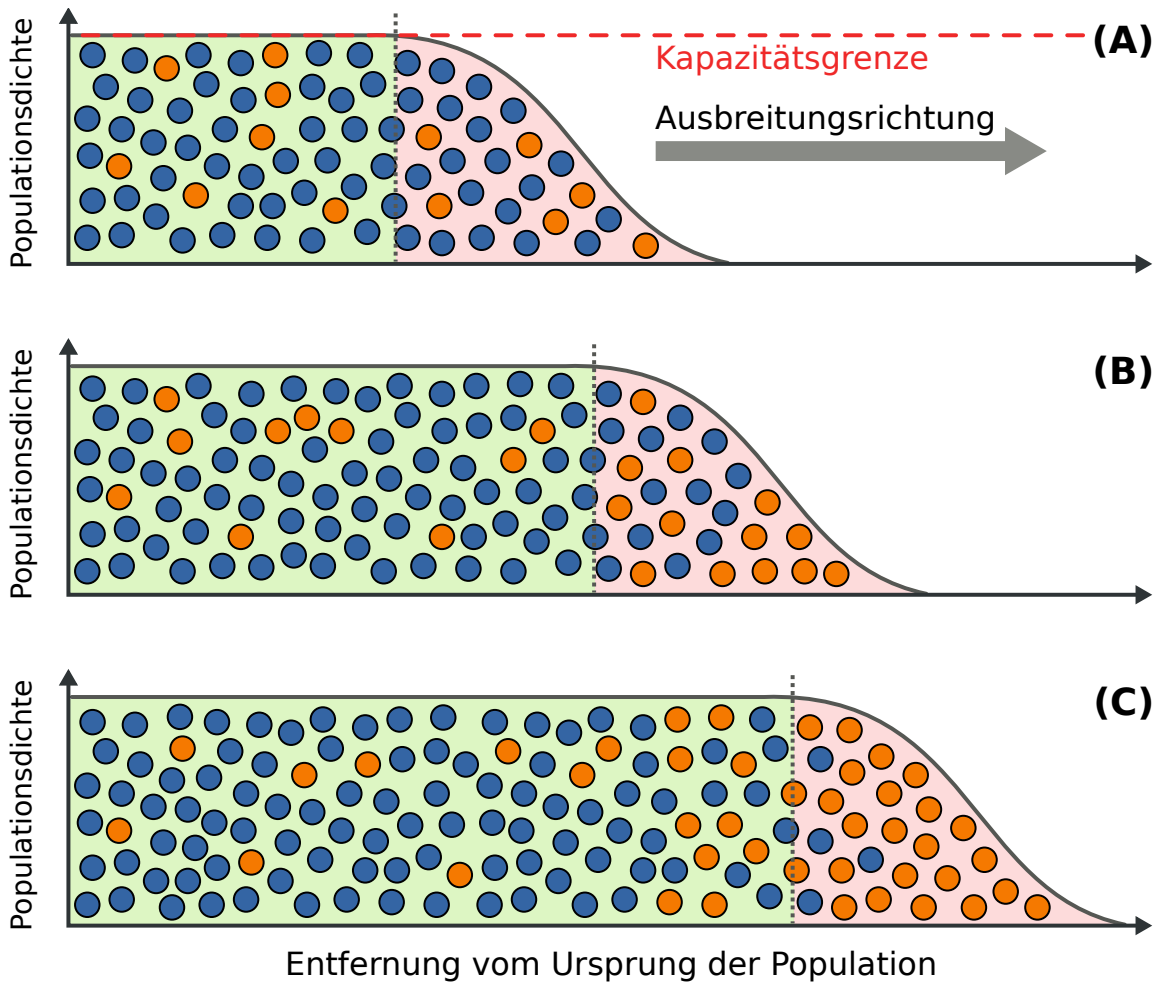


Abbildung 1.3.: **Genetische Drift und Wachstum an der Front einer expandierenden Population.** Dargestellt ist eine expandierende Population, die sich aus haploiden Individuen zweier unterschiedlicher genetischer Varianten (dargestellt als blaue und orangefarbene Kugeln) zusammensetzt. Hinter der Front (grüner Bereich links der gestrichelten Linie) hat die Population die Kapazitätsgrenze erreicht. Hier findet effektiv kein Wachstum statt, da die Individuen im Mittel einen Nachkommen haben. An der Populationsfront (roter Bereich rechts der gestrichelten Linie) ist die Kapazitätsgrenze noch nicht erreicht. Individuen an der Front haben daher im Mittel mehr Nachkommen als Individuen hinter der Front. Die Besiedlung neuer Gebiete findet durch Individuen statt, die zuvor an der Front zu finden waren. Dabei kann es passieren, dass vor allem Individuen einer einzigen genetischen Variante für die Migration ausgewählt werden (Gründereffekt; *founder effect*). Deren Nachkommen sind oftmals an der Besiedlung weiterer Gebiete beteiligt. Findet z.B. die Kolonisation durch ein oder mehrere orangefarbene Individuen statt, so können Träger dieser Variante als Folge fortlaufender Gründereffekte (*serial founder effects*) die Oberhand an der Front gewinnen (B, C). Siehe auch [Excoffier and Ray \(2008\)](#) und [Hallatschek and Nelson \(2008\)](#) (jeweils Abbildung 1).

1. Einleitung

Häufig werden Allelfrequenz-Spektren allein durch statistische Maßzahlen, wie z.B. *Tajimas D*, charakterisiert (Nielsen et al., 2007). Diese Maßzahlen können allerdings keine Aussage darüber treffen, ob Selektion, Wachstum oder eine Range-Expansion für eine bestimmte Verteilung der Allelfrequenzen verantwortlich ist. In dieser Arbeit betrachte ich daher das gesamte Allelfrequenz-Spektrum, um Eigenschaften zu identifizieren, die spezifisch für Range-Expansions sind. Mein Augenmerk richte ich dabei auf die Bereiche hoher Allelfrequenzen, da diese durch das Gene-Surfing stark beeinflusst werden. Wie bereits von (Keinan and Clark, 2012) angeregt werde ich bei meinen Analysen zwischen Mutationen unterscheiden, die bereits in der Ursprungspopulation vorhanden waren und solchen, die während der Range-Expansion entstanden sind.

1.4. Zwei Typen genetischer Variation

Neue genetische Variationen entstehen durch Mutationen. Die Wahrscheinlichkeit, dass von einer Generation zur nächsten an einem Genort des menschlichen Genoms eine Mutation auftritt, ist 1 zu 100 Millionen (1000 Genomes Project Consortium, 2010; Conrad et al., 2011). Wie bereits in ABSCHN. 1.3 beschrieben müssen diese Mutation sehr nahe an der Front auftreten, um vom Surfing-Effekt beeinflusst zu werden (Klopfstein et al., 2006). Bisherige Studien haben daher künstlich einzelne Individuen an einer willkürlichen Stelle der Front mutiert (Edmonds et al., 2004; Klopfstein et al., 2006; Travis et al., 2007). Welche tatsächliche Relevanz fortlaufende Mutationen (*ongoing mutations*, OM) mit realistischen Werten für die Mutationsrate beim Gene-Surfing haben, blieb daher offen (Travis et al., 2007).

In natürlichen Populationen existieren an vielen Genorten genetische Variationen (*standing genetic variation*, SV). So sind für mehr als 36 Millionen der 3.3 Milliarden Genorte des menschlichen Genoms sogenannte SNPs (*single nucleotide polymorphisms*) bekannt (1000 Genomes Project Consortium et al., 2012). Allein zwischen den beiden Chromosomensätzen eines diploiden Individuums gibt es im Schnitt mehr als 3 Millionen Unterschiede (1000 Genomes Project Consortium, 2010). Betrachtet man beide Typen von genetischer Variation in Computersimulationen, so findet man Surfing-Ereignisse sowohl bei Standing-genetic-Variation als auch bei Ongoing-Mutations (ABB. 1.4).

Die Ergebnisse der bereits erwähnten Studie von Keinan and Clark (2012) deuten darauf hin, dass neue Mutationen in menschlichen Populationen den niedrigfrequenten Bereich der Allelfrequenz-Spektren dominieren. Welchen Einfluss Ongoing-Mutations auf

den Bereich hoher Allelfrequenzen haben, ist allerdings unbekannt. Das Vorhandensein von Standing-genetic-Variation legt nahe, dass auch diese Mutationen zu hohen Frequenzen surfen können. Welchen Anteil diese Mutationen aus der Ursprungspopulation am Allelfrequenz-Spektrum nach einer Range-Expansion haben, habe ich in dieser Arbeit untersucht.

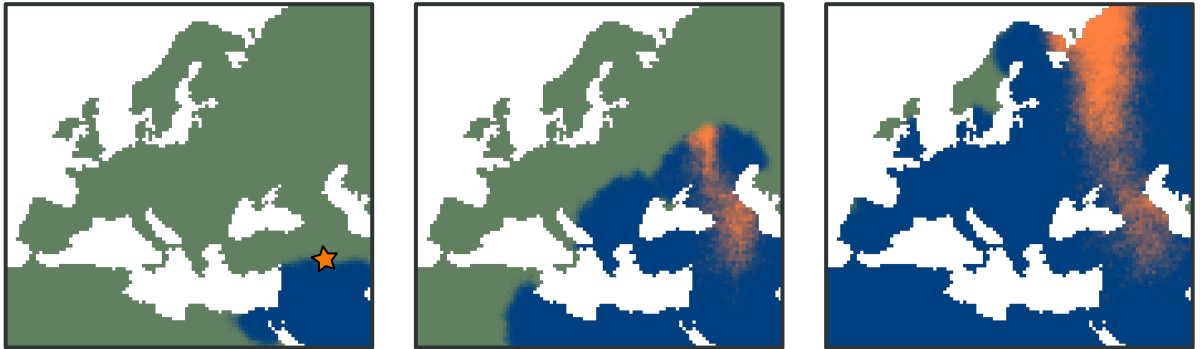
1.5. Ökologische und geographische Faktoren

Ressourcen in der Natur sind oftmals limitiert. Dadurch ist auch die Zahl der Individuen, die an einem Ort existieren können, begrenzt (*Kapazitätsgrenze*). Ein häufig verwendetes Modell, das diese Wachstumsgrenze berücksichtigt, ist das *logistische Wachstum* (Verhulst, 1838, 1845). Beim logistischen Wachstum ist die Pro-Kopf-Zuwachsrate dann am größten, wenn sehr wenige Individuen in der Population vorhanden sind (*negativ-dichteabhängiges Wachstum*). Allerdings fanden Forscher Anfang des letzten Jahrhunderts heraus, dass die Überlebensfähigkeit sehr kleiner Populationen geringer ist, als die größerer Populationen (Allee, 1941). Als Grund dafür wurde die innerartliche Kooperation herangeführt. So kann z.B. eine große Gruppe von Artgenossen einen besseren Schutz vor Feinden bieten (*group protection*). Auch der negative Effekt toxischer Substanzen nimmt mit zunehmender Populationsdichte ab (Allee, 1931, und darin enthaltene Referenzen). Aufgrund der langjährigen Forschung von Warder Clyde Allee in diesem Bereich ist dieses *positiv-dichteabhängige Wachstum* auch als *Allee-Effekt* bekannt.

Weitere Gründe für den Allee-Effekt können die erschwerte Partnersuche sowie ein erhöhtes Maß an Inzucht in kleinen Populationen sein (Courchamp et al., 1999). Letzteres kann zur sogenannten Inzuchtdepression (*inbreeding depression*) führen, die einen negativen Einfluss auf die Überlebens- und Anpassungsfähigkeit einer Population hat (Charlesworth and Willis, 2009; Holsinger and Weir, 2009; Olivieri, 2009). An der Front einer expandierenden Population ist die Populationsdichte gering (ABB. 1.3). In Simulationen von Range-Expansions mit einfachem logistischem Wachstum führt dies sehr schnell zu Bereichen (Sektoren), in denen nur eine einzige genetische Variante zu finden ist (siehe dazu ABB. 1.4). Den Grad an Inzucht, den man in diesen Sektoren erwarten würde, findet man in der Natur allerdings nicht. Daher liegt es nahe, auch solche Wachstumsmodelle in Simulationen zu verwenden, die den Einfluss von Kooperation und Inzucht berücksichtigen.

Neben den Gründereffekten bei der Besiedlung neuer Gebiete können auch geografische

(A) Ongoing-Mutations:



(B) Standing-genetic-Variation:

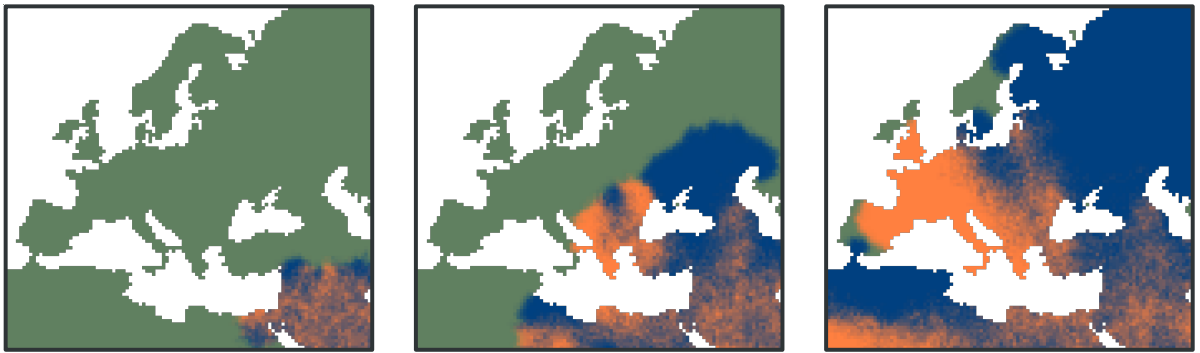


Abbildung 1.4.: **Verteilung genetischer Varianten während der Besiedlung Europas.** Dargestellt sind Simulationen haploider Individuen für einen einzelnen Genort. Am simulierten Genort gibt es zwei neutrale Allele: ein ursprüngliches Allel (blau) und ein abgeleitetes Allel (orange). In (A) tragen zu Beginn alle Individuen der Ausgangspopulation das ursprüngliche Allel. Neue genetische Variation entsteht allein durch Mutationen während der Ausbreitung (*ongoing mutations*). Tritt eine neue Mutation (gekennzeichnet als Stern) nahe der Expansionsfront auf, so kann diese Mutation in einigen der Simulationen eine hohe Frequenz in der Population erreichen. Das ist die Konsequenz wiederholter Gründereffekte an der Expansionsfront und wird als *Gene-Surfing* bezeichnet. In (B) ist bereits vor Beginn der Ausbreitung das abgeleitete Allel (orange) in 50% der Individuen in der Population vorhanden (auch *Standing-genetic-Variation* genannt). Auch hier entstehen während der Range-Expansion Sektoren, die nur eine genetische Variante enthalten. Eine zentrale Fragestellung meiner Arbeit ist, den Unterschied dieser beiden Mutationsmodelle zu quantifizieren. Daher vergleiche ich in dieser Studie den Einfluss von Range-Expansions auf Ongoing-Mutations und Standing-genetic-Variation und bewerte ihren Beitrag zur heutigen genetischen Variation.

Engstellen (*geographic bottlenecks*) während einer Range-Expansion zur Verringerung der genetischen Vielfalt führen. So konnten [Burton and Travis \(2008\)](#) zeigen, dass der Entstehungsort einer Mutation in Relation zu einem geografischen Hindernis einen entscheidenden Einfluss auf die Überlebensfähigkeit dieser Mutation haben kann. Demnach surfen vor allem diejenigen Mutationen, die in der Nähe einer geografischen Engstelle auftreten, während Mutationen abseits der Engstelle diese nur selten passieren können. Da geographische Hindernisse vermutlich bei vielen Range-Expansions eine Rolle gespielt haben, habe ich deren Einfluss auf die Allelfrequenz-Spektren in dieser Arbeit betrachtet.

1.6. Zielsetzung und Struktur dieser Arbeit

In den ersten beiden Abschnitten habe ich dargestellt, dass Range-Expansions ein häufiges Phänomen in der belebten Umwelt sind (ABSCHN. 1.1, ABSCHN. 1.2). Darüber hinaus können Range-Expansions Spuren in genetischen Daten hinterlassen, die denen von selektiven Prozessen und Populationswachstum ähneln (ABSCHN. 1.3). Existierende Methoden können die Spuren von Range-Expansions nicht von den Spuren von Selektion und einfachem Populationswachstum unterscheiden ([Nielsen et al., 2007](#)). Ziel dieser Arbeit ist es daher, den quantitativen Einfluss von Range-Expansions auf genetische Daten zu untersuchen. Dabei widme ich mich vor allem drei Fragen:

1. Gibt es Spuren in genetischen Daten, die spezifisch für Range-Expansions sind?
2. Welche Mutationen werden stärker von Range-Expansions beeinflusst: diejenigen, die bereits in der Ursprungspopulation vorhanden waren, oder solche, die während der Ausbreitung aufgetreten sind?
3. Wie zuverlässig sind meine Ergebnisse unter realen Bedingungen?

Zur Beantwortung dieser Fragen habe ich Computersimulationen durchgeführt, mit denen ich künstliche genetische Daten generiert habe. Das dazu verwendete Modell sowie die Auswertemethoden sind in KAP. 2 erklärt (weitere Details sind in Anhang (KAP. A) zu finden).

Die Ergebnisse in KAP. 3 zeigen, dass Range-Expansions zu Allelfrequenz-Verteilungen führen, die einem Potenzgesetz mit einem spezifischen Exponenten folgen. Im Anhang (KAP. B) präsentiere ich zudem theoretische Überlegungen zum Potenzgesetz sowie einige Ergebnisse aus der Analyse echter genetischer Daten.

1. Einleitung

Meine Ergebnisse werden dazu beitragen, Spuren in echten genetischen Daten korrekt zu interpretieren und Rückschlüsse auf die Vergangenheit zu ziehen (KAP. 4).

2. Methoden

Die populationsgenetischen Prozesse einer sich räumlich ausgebreiteten Population habe ich in Computersimulationen nachgestellt. Ziel dabei war es, künstliche genetische Daten für verschiedene demographische Szenarien zu generieren und anschließend miteinander zu vergleichen. Zum Einstellen der Modellparameter habe ich mich vor allem an Daten menschlicher Populationen orientiert.

Sämtliche Simulationen habe ich auf verschiedenen räumlichen Gittern durchgeführt (siehe TEILABSCHN. 2.1.1). Zur Vereinfachung wurde der geografische Raum in Subpopulationen (*Deme*; Singular: *Dem*) unterteilt. Wie in TEILABSCHN. 2.1.2 beschrieben habe ich die gesamte Dynamik (Reproduktion, Wachstum, Selektion, ...) innerhalb der einzelnen Deme mit einer angepassten Form des Wright-Fisher Modells simuliert (Hartl et al. (2007); Wright (1931); ABB. A.1). Die einzelnen Deme konnten mit einer Migrationsrate m Individuen mit ihren direkten Nachbarn austauschen (*nearest neighbour migration*; siehe ABB. 2.1). Dadurch konnten einzelne Individuen auch bisher unbewohnte Gebiete besiedeln. In neu besiedelten Gebieten wuchs die Population logistisch bis zur Kapazitätsgrenze K (carrying capacity; TEILABSCHN. A.2.2).

In den Simulationen betrachte ich einen einzelnen Genort, an dem sich die Individuen unterschieden (TEILABSCHN. 2.1.3). Entweder hatten sie am simulierten Genort das ursprüngliche (*ancestral*) Allel oder das abgeleitete (*derived*) Allel, das im Folgenden auch als Mutante bezeichnet wird. In getrennten Simulationen habe ich solche Mutationen betrachtet, die entweder mit einer gegebenen Anfangsfrequenz p_s in der Population vorhanden waren (*Standing-genetic-Variation*) oder die mit einer Mutationsrate μ während der Simulation entstehen (*Ongoing-Mutations*). Dabei waren die Mutationen selektiv neutral, so dass ausschließlich genetische Drift für die Änderung der Allelfrequenzen verantwortlich war.

Zunächst habe ich Simulationen dreier demographischer Szenarien miteinander verglichen (siehe TAB. 2.1, TEILABSCHN. 2.1.4). Das Nullmodell war eine Population, die bereits alle Bereiche der Welt besiedelt hatte (*stabile Demographie*). Hier spielte nur

2. Methoden

Tabelle 2.1.: Vergleich demographischer Szenarien. Mit Populationsgröße ist die Größe der Gesamtpopulation gemeint.

Bezeichnung	Räumliche Ausbreitung	Populationsgröße
Stabile Demographie	stabil	stabil
Räumlich stabiles Wachstum	stabil	wachsend (linear)
Range-Expansion	wachsend	wachsend (linear)

die übliche genetische Drift eine Rolle. Die Ergebnisse wurden mit denen einer *Range-Expansion* verglichen, also einer sich räumlich ausbreitenden und wachsenden Population. Um den Einfluss genetischer Drift durch reines Wachstum bewerten zu können, habe ich zusätzlich eine Population mit einem festen Verbreitungsraum simuliert, deren lokale Populationsgröße jedoch gewachsen ist (*räumlich stabiles Wachstum*).

Von Interesse waren dabei vor allem die Frequenzen p_f des abgeleiteten Allels (*derived allele frequencies*, DAF), also des Anteils, den dieses Allel am Genpool der Population hat. Aus den Allelfrequenzen habe ich Allelfrequenz-Spektren erzeugt, in denen die Häufigkeit einer gegebenen Frequenz am Ende der Simulation abgelesen werden kann (TEILABSCHN. 2.2.2). Um auch seltene Ereignisse abbilden zu können, habe ich für jeden Parametersatz in der Regel eine Million Simulationen durchgeführt (TEILABSCHN. 2.2.1).

Das Modell wurde von mir als erweiterbare Bibliothek in *Java SE 6* implementiert (Oracle Corporation, 2011, Version 1.6.0). Als Zufallszahlengenerator innerhalb der Software habe ich *Mersenne Twister* (Matsumoto and Nishimura, 1998) aus der Java Bibliothek *Uncommons Math* (Dyer, 2010, Version 1.2.2) verwendet. Für sämtliche Auswertungen und Plots habe ich die Programmiersprache *R* verwendet (R Core Team, 2013, Version 3.0.2).

Neben den Computersimulationen habe ich auch Allelfrequenz-Spektren für reale menschliche Populationen aus Daten des 1000-Genome-Projektes (1000 Genomes Project Consortium et al., 2012) erstellt und analysiert. Die dazu verwendeten Daten und Methoden beschreibe ich in TEILABSCHN. A.3.2 im Anhang.

2.1. Simulationsmodell

2.1.1. Geographie/Gitter

Ich habe Simulationen auf drei verschiedenen Gittern durchgeführt, die im folgenden als 'Kunstwelt', 'Europakarte' und 'Große Welt' bezeichnet werden (siehe TAB. 2.2). Jeder Punkt des Gitters stellt eine Subpopulation (*Dem*) dar (siehe ABB. 2.1).

Kunstwelt Für die Simulationen der demographischen Szenarien habe ich ein Gitter mit der Breite $w = 25$ und der Höhe $h = 109$ Demen verwendet. Das entspricht in etwa der Größe, die bereits in vorherigen Studien angewendet wurde ($w = 25, h = 100$ in: Edmonds et al., 2004; Klopstein et al., 2006; Travis et al., 2007). Ich habe diesem Gitter 9 weitere Reihen hinzugefügt, da Klopstein et al. (2006) herausgefunden haben, dass in seltenen Fällen auch Individuen aus Demen hinter der Front während einer Range-Expansion vom Gene-Surfing profitieren können. Daher war dieses 'Reservoir' insbesondere bei Simulationen von Standing-genetic-Variation von Bedeutung. Momentaufnahmen einer Simulation auf der Kunstwelt sind in ABB. A.2 gezeigt.

Europakarte Die Simulationen der Besiedlung Europas habe ich auf einer bereits in anderen Studien verwendeten Karte durchgeführt (Klopstein et al., 2006; Currat and Excoffier, 2005). Die Karte war in $110 \times 103 = 11330$ Einheiten unterteilt, wobei jede Einheit einer Fläche von ca. $50km \times 50km = 2500km^2$ entsprach. 4095 dieser Quadrate lagen im Meer, und 12 waren durch das von mir verwendete Nachbarschaftsmodell (Von-Neumann-Nachbarschaft) durch Migration nicht erreichbar. Die übrigen 7223 Quadrate stellten miteinander verbundene Deme dar. Momentaufnahmen einer Simulation auf der Europakarte sind in ABB. A.3 gezeigt.

Große Welt Die Simulationen von Range-Expansions deuteten auf ein Potenzgesetz (*power law*) in den Allelfrequenz-Spektren hin (siehe z.B. ABB. 3.7). Um diese Potenzgesetze für einen großen Frequenzbereich darstellen zu können, habe ich Simulationen mit sehr vielen Individuen auf großen Gittern durchgeführt ($w \times h = 1000 \times 1000$). Zur Verringerung des Rechenaufwandes habe ich die eigentliche Dynamik der Simulation (Reproduktion, Migration) nur in einem begrenzten Bereich um die Front herum simuliert. Das Fenster hatte eine Höhe von $h_{frame} = 45$ Demen und bewegte sich weiter, sobald $\frac{1}{3}$ der möglichen Populationsgröße innerhalb des Fensters erreicht war (siehe auch ABB. A.4(E)). In den Demen außerhalb des

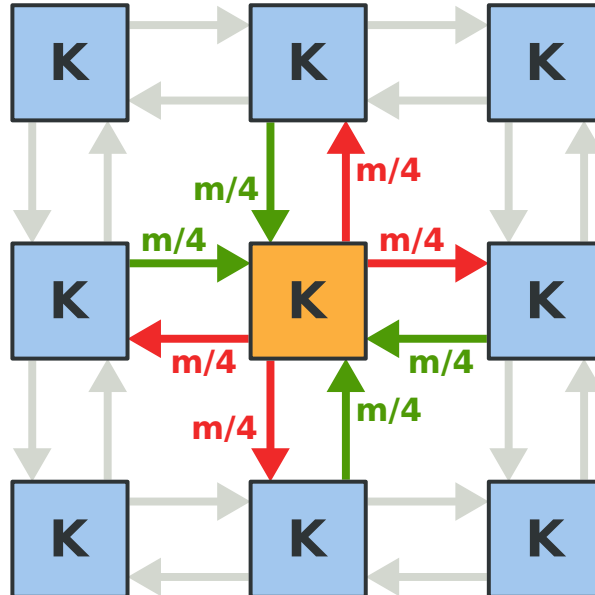


Abbildung 2.1.: **Schematische Darstellung des räumlichen Wright-Fisher Modells.** Gezeigt ist ein Gitter mit 3×3 Subpopulationen (Demen). Innerhalb jedes Dems konnten maximal K Individuen existieren. In Demen mit einer Individuenzahl unterhalb der Kapazitätsgrenze K (*carrying capacity*) wächst die Population logistisch mit der Wachstumsrate r . Die Nachkommen für die nächste Generation wurden für jedes Dem separat mit Hilfe des Wright-Fisher Modells gezogen (siehe ABB. A.1 und ABB. 2.3; Hartl et al. (2007); Wright (1931)). Ein Austausch von Individuen zwischen den Demen erfolgte über Migration (Pfeile). Für das zentrale Dem (orangefarbenes Quadrat) sind Emigration (rote Pfeile) und Immigration (grüne Pfeile) separat dargestellt. In jeder Generation konnte ein Dem mit jedem seiner vier direkten Nachbarn $N_d = \lfloor K \times \frac{m}{4} \rfloor$ Individuen austauschen (GL. 2.7), wobei die Migration simultan durchgeführt wurde (Migrationsrate m). An den Rändern war die effektive Migrationsrate m_e kleiner, da reflektierende Randbedingungen gewählt wurden (GL. 2.8). Die Abfolge der einzelnen Simulationsschritte sind in ABB. 2.2 dargestellt. Innerhalb dieser Arbeit stellt jedes Pixel in Momentaufnahmen einer Simulation genau eine Subpopulation dar.

Tabelle 2.2.: **Übersicht der verwendeten Gitter.** Beispiele für die Gitter sind in ABB. A.2, ABB. A.3 und ABB. A.4 gezeigt. Die Gittergröße ist in Breite mal Höhe ($w \times h$) angegeben. Die Simulationen der demographischen Szenarien in der 'Kunstwelt' unterschieden sich in der Größe des anfänglich besiedelten Bereiches und der Kapazitätsgrenze K . Anzahl anfänglich besiedelter Deme: d_0 .

Gitter	Demographie	Gittergröße	Initialer Bereich	d_0	Demgröße
Kunstwelt	Stabil	25×109	alle Reihen	2725	konstant
Kunstwelt	Wachstum	25×109	alle Reihen	2725	GL. 2.11
Kunstwelt	Expansion	25×109	11 Reihen	275	konstant
Europakarte	Expansion	110×103	ABB. A.3(A)	270	konstant
Große Welt	Expansion	1000×1000	11 Reihen	11'000	konstant

Fensters blieb die zuvor entstandene Zusammensetzung der Population bis zum Ende der Simulation erhalten. Drift-Effekte abseits der Front spielten hier somit keine Rolle. Momentaufnahmen einer Simulation auf der Großen Welt sind in ABB. A.4 gezeigt. Zudem habe ich Simulationen auf einer Großen Welt mit geänderten Proportionen ($w \times h = 10000 \times 100$) sowie geänderten Größen ($w \times h = 500 \times 500$ bzw. $w \times h = 5000 \times 50$) durchgeführt. Dadurch konnte ich die Auswirkungen der eigentlichen Lauflänge sowie der Frontbreite ermitteln.

2.1.2. Eigenschaften der Deme

In jedem Dem des Gitters (ABB. 2.1) befand sich eine angepasste Wright-Fisher Population (ABB. 2.3), die bis zu einer Kapazitätsgrenze (*carrying capacity*) von K Individuen wachsen konnte. Jeder Zeitschritt der Simulation (*Generation*) bestand aus einer Reproduktions- und einer Migrationsphase, die nachfolgend beschrieben werden (siehe auch ABB. 2.2).

Reproduktion, Wachstum und Allee-Effekt Im klassischen *Wright-Fisher Modell für genetische Drift* (ABB. A.1; Hartl et al. (2007); Wright (1931)) besteht die Population aus Individuen mit zwei verschiedenen neutralen genetischen Varianten, die gemeinsam alle K Plätze der Population belegen. Die Zusammensetzung der Population in der nächsten Generation wird durch K -maliges Ziehen aus einer Binomialverteilung der Form

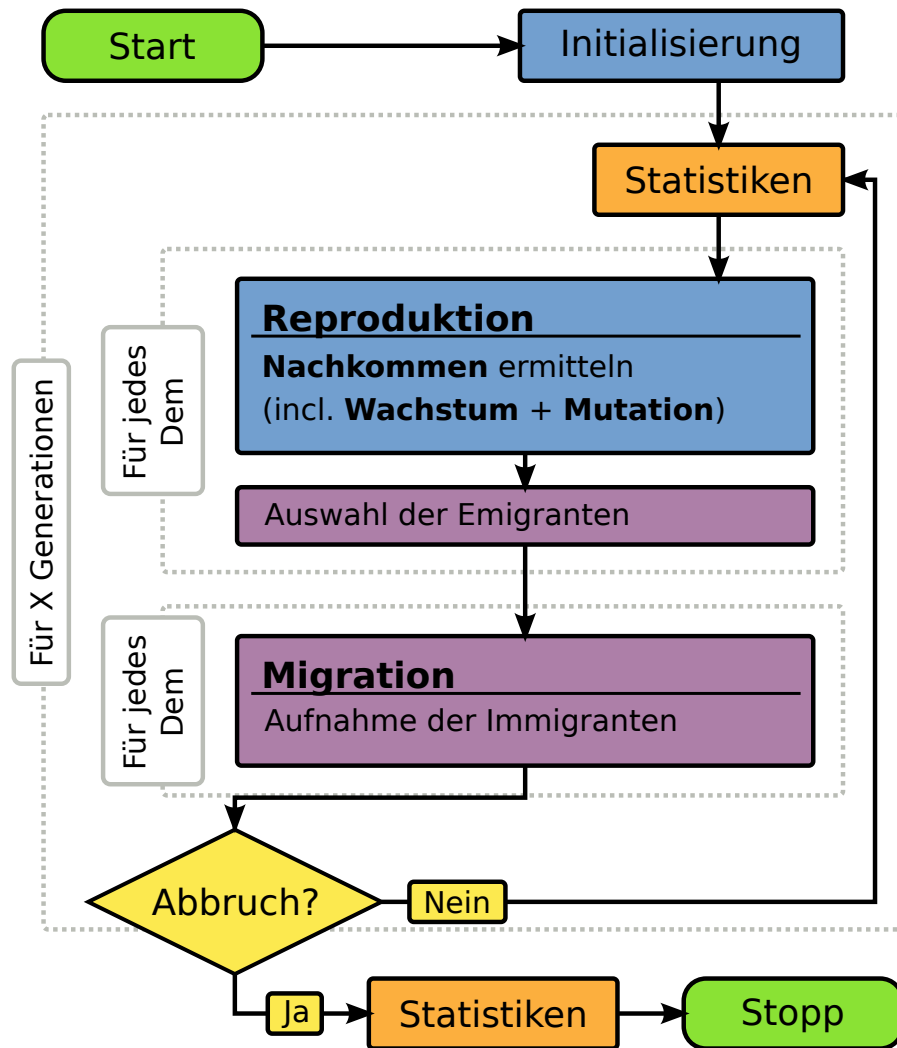


Abbildung 2.2.: **Flussdiagramm des Simulationsablaufs.** *Blaue Felder* beinhalten Prozesse, bei denen die Zusammensetzung der Population durch das Ziehen aus einer Multinomialverteilung ermittelt wird. Während der Initialisierung werden (abhängig vom demographischen Szenario; TAB. 2.2) einzelne Deme mit Individuen besetzt. In der Reproduktionsphase wird die Zusammensetzung der aktuellen Population aus der Eltern-Generation ermittelt. In diesem Schritt kann die Population wachsen und es können Mutationen auftreten. *Violette Felder* beinhalten Schritte der Migration. Zunächst werden die Emigranten multi-hypergeometrisch gezogen und im Anschluss ausgetauscht. Die Prozesse, die innerhalb der Deme stattfinden (*gepunktete Kästchen* in der Mitte), sind in TEILABSCHN. 2.1.2 und TEILABSCHN. 2.1.3 beschrieben.

$$B(N_1|p_1, K) = \binom{K}{N_1} \cdot p_1^{N_1} \cdot \underbrace{(1 - p_1)^{K - N_1}}_{\cong p_2} \quad (2.1)$$

mit

$$\begin{aligned} N_1 &= 0, 1, \dots, K \\ p_1 + p_2 &= 1 \end{aligned}$$

ermittelt. Dabei ist N_1 die Anzahl der Nachkommen von Individuen der Variante 1, $p_1 \equiv \frac{N_1}{K}$ ist die Frequenz dieser Variante, $p_2 = (1 - p_1)$ ist die Frequenz von Individuen der Variante 2, und K ist die Kapazitätsgrenze. Die tatsächliche Zusammensetzung der Population ändert sich in der Regel von einer Generation zur nächsten. Die erwarteten Frequenzen der Varianten in der Nachkommen-Generation entsprechen in diesem Modell immer den Frequenzen der aktuellen Generation:

$$\begin{aligned} \mathbb{E}(p'_1) &= p_1 &= \frac{N_1}{K} \\ \mathbb{E}(p'_2) &= p_2 = (1 - p_1) &= \frac{N_2}{K}. \end{aligned} \quad (2.2)$$

Im Gegensatz zum klassischen Wright-Fisher Modell sind bei Simulationen von Range-Expansions nicht alle Deme bis zur Kapazitätsgrenze gefüllt. Um das Wachstum innerhalb derjenigen Deme zu simulieren, die sich unterhalb der Kapazitätsgrenze befanden, habe ich eine weitere Variante in das Modell eingeführt, welche *leere Plätze* repräsentiert. Wie die beiden Typen von *Individuen* (*genetischen Varianten*; p_1 und p_2) werden auch die leeren Plätze in diesem Modell durch eine Frequenz (p_0) repräsentiert. Das Wachstum kommt dadurch zustande, dass die leeren Plätze einen selektiven Nachteil gegenüber den genetischen Varianten haben (ABB. 2.3). Dadurch sinkt die Wahrscheinlichkeit stetig, leere Plätze in der nächsten Generation zu finden. Dieser Einfluss von Selektion muss bei der Berechnung der erwarteten Wahrscheinlichkeiten mit einbezogen werden. Dazu habe ich die frequenzabhängige Selektion (Replikator-Gleichung; Nowak (2006)) verwendet, durch die logistisches Wachstum simuliert wird (siehe dazu TEILABSCHN. A.2.2). Die erwartete Wahrscheinlichkeit $\mathbb{E}(p'_i)$ der Variante i (von n möglichen Varianten) in der nächsten

2. Methoden

Generation ergibt sich aus

$$\mathbb{E}(p'_i) = p_i + \sum_{\substack{j=0 \\ j \neq i}}^n \underbrace{p_i \cdot p_j \cdot s_{i,j}}_{\substack{\text{Änderung durch} \\ \text{Konkurrenz} \\ \text{mit Variante } j}} \quad (2.3)$$

mit

$$\begin{aligned} \sum_{i=0}^n \mathbb{E}(p'_i) &= 1 \\ \sum_{i=0}^n \sum_{j=0}^n s_{i,j} &= 0. \end{aligned}$$

Dabei ist $s_{i,j}$ die Differenz zwischen dem Selektionskoeffizienten der Variante i und dem Selektionskoeffizienten der Variante j ($s_{i,j} = s_i - s_j$). Die Werte für die Selektionskoeffizienten waren dabei so gewählt, dass $s_{i,j}$ zwischen zwei neutralen genetischen Varianten 0 war und zwischen einer genetischen Variante und den leeren Plätzen genau der Pro-Kopf-Wachstumsrate r entsprach (GL. 2.4). Bei einer Wachstumsrate r findet man pro Individuum im Mittel $r + 1$ Nachkommen in der Folgegeneration. In den meisten Simulationen war $r = 0.5$ (TAB. 2.3). Das liegt im Bereich der Wachstumsrate für menschliche Populationen, die mit Werten zwischen $r = 0.1$ und $r = 0.8$ angegeben wird (Coventry et al., 2010; Klopstein et al., 2006; Currat and Excoffier, 2005; Wegmann et al., 2006; Edmonds et al., 2004; Ray and Excoffier, 2010).

$$\begin{bmatrix} s_{0,0} & s_{0,1} & \dots & s_{0,n} \\ s_{1,0} & s_{1,1} & \dots & s_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n,0} & s_{n,1} & \dots & s_{n,n} \end{bmatrix} = \begin{bmatrix} 0 & -r & \dots & -r \\ +r & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ +r & 0 & \dots & 0 \end{bmatrix} \quad (2.4)$$

Nach der Errechnung der erwarteten Frequenzen $\{\mathbb{E}(p'_i)\}_{0 \leq i < n}$ für jede der n Varianten wurde die tatsächliche Zusammensetzung der Folgegeneration (für jedes Dem) durch Ziehen aus einer Multinomialverteilung mit diesen Parametern ermittelt (siehe CODE A.1; Devroye (1986)).

Beim logistischen Wachstum ist die Pro-Kopf-Wachstumsrate dann am höch-

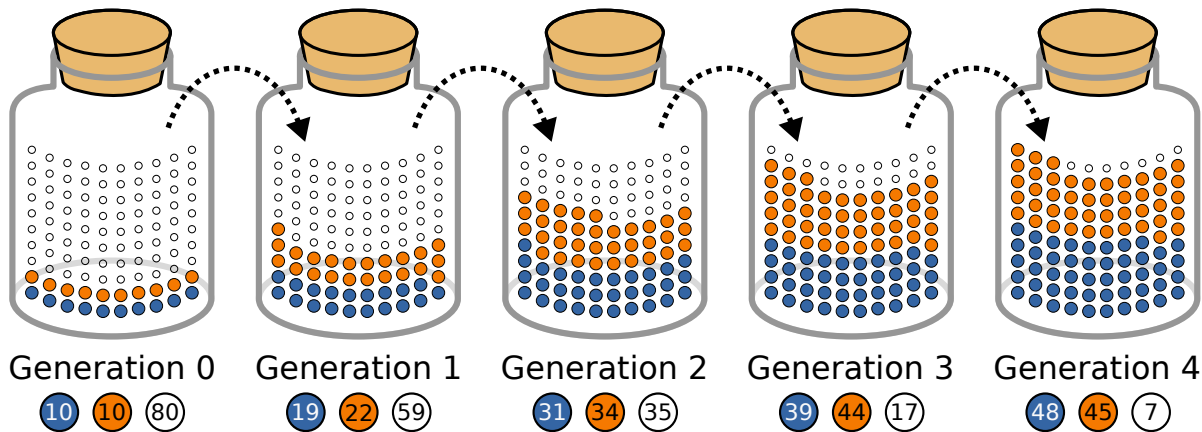


Abbildung 2.3.: **Angepasstes Wright-Fisher Modell für die Simulation des Wachstums innerhalb nicht komplett besiedelter Deme.** Wie in [ABB. A.1](#) sind die Individuen durch verschiedenfarbige Kugeln dargestellt. Während orangefarbene und blaue Kugeln tatsächliche Individuen darstellen, repräsentieren weiße Kugeln die noch nicht besetzten Bereiche der Population bzw. des Dems. Zur Ermittlung der nächsten Generation (gepunktete Pfeile) werden die weißen Kugeln wie die farbigen Kugeln gezogen, allerdings mit einer geringeren Wahrscheinlichkeit (hier symbolisiert durch die geringere Größe der weißen Kugeln). Dadurch kommt es zu einem Wachstum (logistische Pro-Kopf-Wachstumsrate: $r = 1.0$; Kapazitätsgrenze: $K = 100$; Zahl der Individuen in der Generation 0: $N_0 = 20$). Der hier dargestellte Prozess entspricht dem Ziehen aus einer Multinomialverteilung. Ein Austausch von Individuen mit benachbarten Demen ist in diesem Beispiel nicht dargestellt. Bildvorlage: [KAP. C](#).

Tabelle 2.3.: Typische Parameter der Simulationen.

Parameter	Typischer Wert	Bedeutung
r	0.5	<i>Wachstumsrate</i> . Ein Wert von $r = 1$ bewirkt, dass sich die Population von einer Generation zur nächsten im Mittel verdoppelt. Ein Wert von $r = 0.5$ bewirkt, dass jedes Individuum im Schnitt 1.5 Nachkommen zur nächsten Generation beiträgt. Beim logistischen Wachstum werden diese Zuwachsraten nur bei geringer Populationsdichte erreicht (ABB. A.5).
K	100	<i>Kapazitätsgrenze</i> (engl. carrying capacity). Maximale Anzahl von Individuen, die innerhalb eines Dems existieren können.
m	0.2	<i>Migrationsrate</i> . Anteil der Individuen, die in jeder Generation das Dem verlassen — hier: 20%ige Wahrscheinlichkeit eines Individuums, als Migrant gezogen zu werden.
μ	10^{-6}	<i>Mutationsrate</i> bei Simulationen von Ongoing-Mutations. Die Mutationsrate ist die Wahrscheinlichkeit, dass an einem Genort von einer Generation zur nächsten eine Mutation auftritt. Für den Menschen wurde eine Mutationsrate von 10^{-8} ermittelt (Conrad et al., 2011).
p_s	10^{-3}	<i>Startfrequenz (initial mutant frequency)</i> bei Simulationen von Standing-genetic-Variation. Wahrscheinlichkeit eines Individuums, Träger des abgeleiteten Allels zu sein (hier: 1 von 1000 Individuen). Bereich simulierter Startfrequenzen: $p_s = 10^{-5}$ bis $p_s = 0.5$.
α	<i>n.a.</i>	Bei Simulationen des Allee-Effektes gibt α die Position der halb-maximalen Wachstumsrate an. Ein Wert von $\alpha = 0.1$ bedeutet, dass bei einer Populationsdichte von 10% die Wachstumsrate genau $0.5 \times r$ beträgt (siehe ABB. A.6). Ist der Wert nicht angegeben ($\alpha = n.a.$), gibt es keinen Allee-Effekt.
ϵ	<i>n.a.</i>	Bei Simulationen des Allee-Effektes bestimmt ϵ die Breite des Bereiches, in dem sich Wachstumsrate von 0 auf r ändert. Ein hoher ϵ -Wert führt dazu, dass die Wachstumsrate bei einer Populationsdichte um α sehr schnell von 0 auf r ansteigt (ABB. A.6). Ist der Wert nicht angegeben ($\epsilon = n.a.$), gibt es keinen Allee-Effekt.

ten, wenn sehr wenige Individuen in der Population vorhanden sind (negativ-dichteabhängiges Wachstum; ABB. A.5). In sich sexuell fortpflanzenden Populationen kann eine niedrige Populationsdichte allerdings die Partnersuche erschweren oder die Überlebensfähigkeit der Nachkommen aufgrund von Inzuchtdepression herabgesetzt sein. Dadurch verringert sich die Pro-Kopf-Wachstumsrate r bei sinkender Populationsdichte (*positiv-dichteabhängiges Wachstum*; Allee-Effekt (Allee, 1941; Courchamp et al., 1999)). Um diesen Allee-Effekt zu simulieren, habe ich einen Faktor AF eingeführt, der abhängig von der aktuellen Populationsdichte die Pro-Kopf-Wachstumsrate r herabsetzt. Dadurch ändert sich die Wahrscheinlichkeit, leere Felder in der nächsten Generation zu ziehen (ABB. 2.4). Abgeleitet von GL. 2.3 ergibt sich

$$\mathbb{E}(p'_i) = p_i + \sum_{\substack{j=0 \\ j \neq i}}^n p_i \cdot p_j \cdot \underbrace{s_{i,j} \cdot AF}_{\substack{\text{positiv-} \\ \text{dichteabhängige} \\ \text{Wachstumsrate}}} \quad (2.5)$$

mit

$$AF = \begin{cases} 1 & \text{wenn } i, j \neq 0, \\ \frac{(N/K)^\epsilon}{(N/K)^\epsilon + \alpha^\epsilon} = \frac{(1-p_0)^\epsilon}{(1-p_0)^\epsilon + \alpha^\epsilon} & \text{sonst.} \end{cases} \quad (2.6)$$

Ich habe einen Allee-Effekt simuliert, bei dem es kein negatives Wachstum bei geringen Populationsdichten gibt (*nicht-kritischer Allee-Effekt*; Taylor and Hastings (2005)). In GL. 2.6 gibt $N = (\sum_{i=1}^n N_i)$ die Zahl der nicht leeren Plätze an. Der Faktor α entspricht derjenigen Populationsdichte, bei der das Wachstum halbmaximal ist. Der Faktor ϵ bestimmt, wie stark das Wachstum der Population vor dem Erreichen der Populationsdichte α ist und wie schnell die maximale Pro-Kopf-Wachstumsrate r erreicht wird (siehe dazu ABB. A.6 und TAB. 2.3). Sofern α oder ϵ nicht angegeben (*n.a.*) sind, wird kein Allee-Effekt simuliert ($AF = 1$).

Migration und Randbedingungen In der Migrationsphase tauschte jedes Dem mit seinen vier direkten Nachbarn Individuen aus (ABB. 2.1; Von-Neumann-Nachbarschaft). Diese Migration über kurze Distanzen entspricht laut aktuellen Studien dem Migrationsverhalten der Menschen bei der Besiedlung der Welt (Eswaran, 2002). So konnten z.B. Prugnolle et al. (2005) zeigen, dass die genetische Vielfalt linear

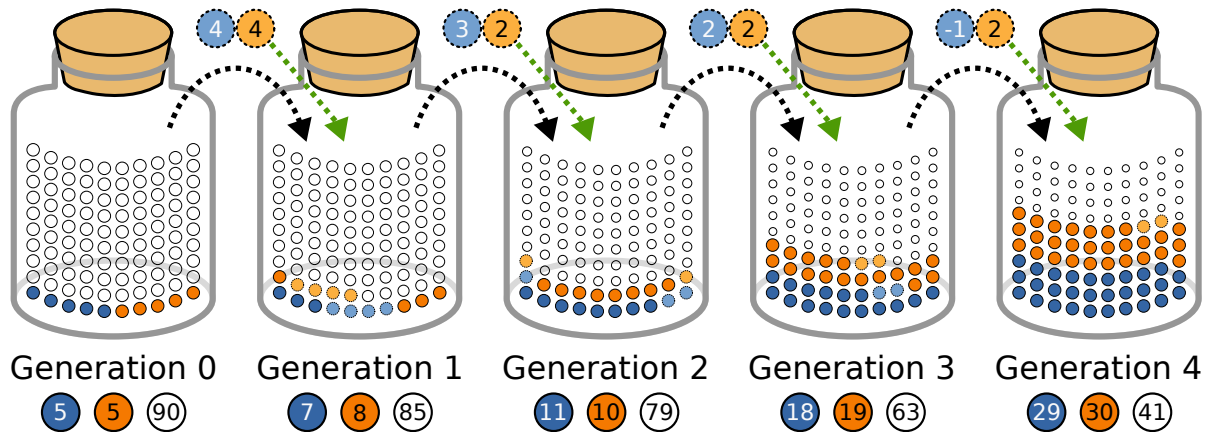


Abbildung 2.4.: **Angepasstes Wright-Fisher Modell für die Simulation des Allee-Effektes.** Wie in ABB. 2.3 sind tatsächliche Individuen durch blaue bzw. orangefarbene und „leere Plätze“ durch transparente Kugeln dargestellt. Die Zusammensetzung der nächsten Generation wird auch hier zunächst durch Ziehen aus einer Binomialverteilung ermittelt (schwarze Pfeile). Der Allee-Effekt führt dazu, dass in dünnbesiedelten Demen das Wachstum verringert ist. Daher sind zu Beginn die weißen Kugeln gleich groß dargestellt. Darüber hinaus wird in diesem Beispiel die Netto-Immigration von Individuen aus benachbarten (dichter besiedelten) Demen gezeigt (grüne Pfeile). Die Zahlen in den Kreisen mit gestrichelten Rändern geben an, wie viele Individuen des entsprechenden Typs mehr ein- als auswandern. Die Migration ermöglicht es, die kritische Populationsdichte α zu erreichen, ab der die Population wachsen kann. Bildquelle: KAP. C. ($r = 1.0$, $\alpha = 0.25$)

mit der Distanz von Afrika abnimmt und die genetische Differenzierung zweier Populationen mit der Distanz zunimmt. Sie führten dies auf aufeinanderfolgende Gründereffekte (*serial founder effects*) sowie begrenzten Genfluss (*isolation by distance*) zurück.

In meinem Modell findet die Migration direkt im Anschluss an die Reproduktionsphase statt (ABB. 2.2). Dazu wurden Individuen eines Dems mit der Wahrscheinlichkeit m (Migrationsrate) ausgewählt und gleichmäßig auf die $d = 4$ möglichen Richtungen aufgeteilt. Die Zahl N_d der Migranten ($\lfloor \]$ bedeutet „ganzzahlig abgerundet“), die mit einem gegebenen Nachbarn ausgetauscht wurden, war:

$$N_d = \left\lfloor K \cdot \frac{m}{4} \right\rfloor \quad (2.7)$$

In den meisten Simulationen war $m = 0,2$ (TAB. 2.3). Das liegt im Bereich der Migrationsraten, die in anderen Studien für die Simulation der Ausbreitung des Menschen verwendet wurden ($m = 0.05$ bis $m = 0.4$) (Wegmann et al., 2006; Ray and Excoffier, 2010; Klopstein et al., 2006; Currat and Excoffier, 2005; Edmonds et al., 2004).

Die Ränder des Gitters und die geographischen Hindernisse (z.B. Meere) waren reflektierend (*reflecting boundary condition*). Individuen, die in diese Richtungen geschickt werden sollten, verblieben in ihrem ursprünglichen Dem. Durch dieses Verhalten treten keine „Randeffekte“ (*boundary effects*) auf, wie sie von Burton and Travis (2008) beschrieben wurden, sondern die Front bleibt im Mittel gerade. Die Zahl der Nachbarn d in den Randbereichen lag in der Kunstwelt bei $d = \{2, 3, 4\}$ und auf der Europakarte bei $d = \{1, 2, 3, 4\}$. Die effektive Migrationsrate m_e eines Dems war demzufolge:

$$m_e = m \cdot \frac{d}{4}. \quad (2.8)$$

Bei der Auswahl der Migranten wurden leere Felder wie genetische Varianten behandelt, so dass durch einen Austausch von leeren Feldern und tatsächlichen Individuen die Neubesiedlung möglich wird. Um Individuen nicht mehrfach für die Migration zu ziehen, habe ich anstelle der Multinomialen Verteilung (‘Ziehen *mit* Zurücklegen’) eine Multi-Hypergeometrische Verteilung (‘Ziehen *ohne* Zurücklegen’) verwendet (Devroye (1986); CODE A.2).

2.1.3. Arten und Verteilung der genetischen Variation

Die Individuen in den Simulationen unterschieden sich an einem einzelnen Genort, an dem es $n - 1$ genetische Varianten (Allele) gab. In der Regel habe ich Simulationen für $n = 3$ Varianten durchgeführt; 0: leerer Platz; 1: ursprüngliches Allel (Wildtyp); 2: abgeleitetes Allel (Mutante). Die genetischen Varianten waren paarweise selektiv neutral, so dass keine der beiden einen Wachstumsvorteil gegenüber der jeweils anderen hatte (siehe auch GL. 2.4). Mutationen waren entweder zu Beginn der Simulation vorhanden (*Standing-genetic-Variation*) oder sie entstanden während der Simulation durch Mutation (*Ongoing-Mutations*).

Standing-genetic-Variation Bei der Analyse von *Standing-genetic-Variation* wurde zu Beginn der Simulation die Zusammensetzung jedes Deme innerhalb des initialen Bereiches (TAB. 2.2) durch Ziehen aus einer Multinomialverteilung ermittelt (ABB. 2.2). Da die Deme im initialen Bereich voll besetzt waren, war die erwartete Frequenz von leeren Plätzen $p_0 = 0$. Die erwarteten Frequenzen der Mutation und des Wildtyps waren $p_2 = p_s$ bzw. $p_1 = 1 - p_s$. Die Deme außerhalb des initialen Bereiches wurden vollständig mit leeren Plätzen belegt ($p_0 = 1$). Für die Anfangsfrequenz p_s der Mutation habe ich in den Simulationen Werte von $p_s = 10^{-5}$ bis $p_s = 0,5$ verwendet (TAB. 2.3).

In Simulationen der großen Welt gab es 124 verschiedene Allele ($n = 125$), jeweils mit einer Anfangsfrequenz von $p_s = \frac{1}{124} \approx 0.00865$. Die gleiche Anfangsfrequenz der Allele ermöglichte es mir, 124 Simulationen in einer zusammenzufassen und dadurch den Rechenaufwand zu reduzieren. In der anschließenden Analyse habe ich jede der genetischen Varianten einmal als Mutante und den Rest als Wildtyp betrachtet.

Ongoing-Mutations In Simulationen von *Ongoing-Mutations* können in jeder Generation Mutationen mit einer Wahrscheinlichkeit von μ (Mutationsrate) am betrachteten Genort auftreten. Diese Mutationen ändern die erwarteten Frequenzen $\mathbb{E}(p'_i)$ einer genetischen Variante i in der nächsten Generation (GL. 2.9). Die erwartete Frequenz wird größer, wenn eine entsprechende Mutation in Individuen einer anderen genetischen Variante (j) auftritt (*Zugewinn*). Mutieren allerdings Individuen der Variante i , so sinkt ihre erwartete Wahrscheinlichkeit (*Verlust*). Welche Individuen tatsächlich mutieren, ist dabei von der aktuellen Frequenz p_i bzw. p_j abhängig. Besteht z.B. eine Population fast ausschließlich aus Individuen der Variante i , so ist

es wahrscheinlicher, dass Individuen dieser Variante von den Mutationen betroffen sind. Abgeleitet von GL. 2.3 ergibt sich:

$$\mathbb{E}(p'_i) = p_i + \sum_{\substack{j=0 \\ j \neq i}}^n \left(\overbrace{p_i \cdot p_j \cdot s_{i,j}}^{\text{Änderungen durch Selektion/Wachstum}} + \underbrace{(\mu_{j,i} \cdot p_j)}_{\text{Zugewinn}} - \underbrace{(\mu_{i,j} \cdot p_i)}_{\text{Verlust}} \right). \quad (2.9)$$

Dabei sind $\mu_{i,j}$ und $\mu_{j,i}$ die Wahrscheinlichkeiten, dass ein Allel der Variante i zu einem Allel der Variante j bzw. ein Allel der Variante j zur Variante i mutiert. In Simulationen mit zwei genetischen Varianten entsprechen diese Übergangswahrscheinlichkeiten genau der Mutationsrate μ . Eine Umwandlung von leeren Plätzen zu genetischen Varianten oder von genetischen Varianten zu leeren Plätzen ist nicht möglich. Daher ergibt sich folgende Mutationsmatrix:

$$\begin{bmatrix} \mu_{0,0} & \mu_{0,1} & \mu_{0,2} \\ \mu_{1,0} & \mu_{1,1} & \mu_{1,2} \\ \mu_{2,0} & \mu_{2,1} & \mu_{2,2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \mu \\ 0 & \mu & 0 \end{bmatrix}. \quad (2.10)$$

Die simulierten Mutationsraten lagen im Bereich von $\mu = 10^{-5}$ und $\mu = 10^{-7}$ und damit ein bis drei Größenordnungen oberhalb der für den Menschen ermittelten Mutationsrate von $\mu \approx 10^{-8}$ (1000 Genomes Project Consortium, 2010; Conrad et al., 2011; Coventry et al., 2010). Allerdings zeigten meine Simulationsergebnisse, dass die resultierenden Allelfrequenz-Spektren um die entsprechenden Größenordnung entlang der Abszisse verschoben sind (ABB. 3.3 und ABB. B.4).

2.1.4. Demographische Szenarien

Ich habe drei verschiedene demographische Szenarien simuliert: 'Stabile Demographie', 'Range-Expansion' und 'Räumlich stabiles Wachstum' (TAB. 2.1). Die Szenarien unterscheiden sich zum einen in der Größe des initialen Bereiches, zum anderen in der Kapazitätsgrenze der Deme (siehe TAB. 2.2).

Stabile Demographie Bei der stabilen Demographie sind alle Deme von Beginn an bis zur Kapazitätsgrenze K mit Individuen besiedelt. Es gibt keine leeren Plätze,

2. Methoden

so dass weder eine räumliche Ausbreitung, noch ein Wachstum der Population stattfinden kann.

Range-Expansion Bei Range-Expansions sind nur die Deme innerhalb eines „initialen Bereiches“ mit Individuen besiedelt (TAB. 2.2). Die restlichen Deme waren der „neu besiedelte Bereich“ und enthielten zu Beginn einer Simulation ausschließlich leere Plätze ($p_0 = 1$). In den Simulationen auf der Kunstwelt und der großen Welt dienten die ersten 11 Reihen als initialer Bereich (ABB. A.2(A), ABB. A.4(A)). Im Gegensatz zu Klopstein et al. (2006) habe ich anstelle eines einzelnen zentralen Deme die kompletten Reihen besiedelt. Mit der so entstehenden geraden Front sollten eventuelle Effekte einer gebogenen (und dadurch längeren) Front so gering wie möglich gehalten werden. Bei Simulationen auf der Europakarte deckte der initiale Bereich die angenommenen Startpunkte der Paläolithischen und der Neolithischen Besiedlung ab (ABB. A.3(A); Currat and Excoffier, 2005, Abbildung 1).

Räumlich stabiles Wachstum Wie bei der stabilen Demographie sind auch beim räumlich stabilen Wachstum zu Beginn der Simulation alle Deme bis zur Kapazitätsgrenze besiedelt. Dadurch findet keine räumliche Ausbreitung statt. Die Population wächst allerdings, da in diesem Szenario die aktuelle Kapazitätsgrenze $K_c(t)$ der Deme mit der Zeit t von der anfänglichen Kapazitätsgrenze K_s bis zu einer maximalen Kapazität K ansteigt.

$$K_c(t) = \min(K, \underbrace{\lfloor r_{deme} \cdot t + K_s \rfloor}_{\text{linearer Anstieg}}) \quad (2.11)$$

Zu jedem Zeitpunkt der Simulation war die Gesamtgröße der Population vergleichbar mit derjenigen einer Range-Expansion. Da das Wachstum der Gesamtpopulation bei einer *Range-Expansion* näherungsweise linear ist (siehe ABB. B.1), kann der Anstieg der Kapazitätsgrenzen durch eine Geradengleichung beschrieben werden. Die Steigung r_{deme} habe ich durch lineare Regression aus 10^6 Range-Expansion-Simulationen ermittelt (siehe ABB. B.1). Die Anzahl der Individuen zu Beginn der Simulation entsprach der Größe der Ursprungspopulation einer Range-Expansion ($N_s = K \times w \times 11 = 27500$). Der Abszissenabschnitt K_s , der sich aus der Anzahl der Individuen zu Beginn der Simulation, geteilt durch die Anzahl der Deme, errechnet, war daher immer $K_s = 10.19$.

2.2. Umsetzung des Modells

2.2.1. Laufzeiten und Wiederholungen

Die Simulationen auf der Kunstwelt und auf der Europakarte liefen für jeweils $g = 2000$ Generationen. Bei einer Generationszeit des Menschen von etwa 25 – 30 Jahren entspricht das einem Zeitraum von 50'000 – 60'000 Jahren (Langergraber et al., 2012, sowie die darin enthaltenen Referenzen). Die Range-Expansions dauerten (abhängig von den Parametern) auf der Kunstwelt ca. $t_f = 370$ Generationen (TAB. 3.2) und auf der Europakarte ca. $t_f = 420$ Generationen (TAB. B.1). Für jeden Parametersatz wurden die Simulationen mindestens 10^6 mal durchgeführt.

Die Simulationen auf der Großen Welt wurden beendet, sobald ein Individuum die letzte Reihe des Gitters erreicht hat (siehe auch ABB. A.4(F) und TAB. 3.3). Dadurch trat nach dem eigentlichen Ende der Range-Expansion keine genetische Drift mehr auf. Für jeden Parametersatz wurden auf der Großen Welt 800 Simulationen mit je $n = 124$ genetischen Varianten durchgeführt. Dadurch konnten im Anschluss 99200 Frequenzen abgeleiteter Allele gemessen werden.

2.2.2. Messgrößen

In den Simulationen habe ich zu verschiedenen Zeitpunkten die Frequenz p_f des abgeleiteten Allels gemessen. In Simulationen von Range-Expansions habe ich dazu die Deme des *neu besiedelten Bereiches* herangezogen. In Simulationen der stabilen Demographie und dem räumlich stabilen Wachstum waren bereits von Beginn an alle Deme besiedelt. Daher habe ich bei diesen demographischen Szenarien alle Deme des Gitters für die Messung verwendet.

Aus den Frequenzen aller Simulationen für einen Parametersatz habe ich Dichtefunktionen der Allelfrequenz-Spektren sowie kumulierte Häufigkeitsverteilungen (Zipf-Plots; Newman (2005)) erstellt. Details dazu stehen in TEILABSCHN. A.3.1). Zur Abschätzung der Fehler in den Allelfrequenz-Spektren habe ich ein Bootstrap-Verfahren angewendet, das ebenfalls in TEILABSCHN. A.3.1 beschreiben ist.

Aus den Allelfrequenz-Spektren für Standing-genetic-Variation (SV) und Ongoing-Mutations (OM) konnte ich weitere Spektren ($\Gamma(p_f)$) erstellen, mit deren Hilfe ich Aussagen über den relativen Beitrag von SV und OM zu den Allelfrequenz-Spektren

Tabelle 2.4.: Primäre Messgrößen der Simulationen.

Messgröße	Bedeutung
p_f	(Finale) Allelfrequenz. Anteil des abgeleiteten Allels am Genpool der Population zum Ende der Simulation (bzw. zu einem bestimmten Zeitpunkt). Aus diesen Werten habe ich die Allelfrequenz-Spektren erstellt.
Γ	Γ ist ein Wert, der Aussagen über die relative Bedeutung von Standing-genetic-Variation und Ongoing-Mutations für die Allelfrequenz-Spektren nach einer Range-Expansion erlaubt (siehe dazu TAB. 3.1). Gemessen wurde dieser Wert in einem Bereich, in dem die Range-Expansion für gleiche Steigungen in den Allelfrequenz-Spektren von SV und OM verantwortlich war (siehe ABB. 3.4). Ergebnisse in TAB. 3.2.
a_k	Exponent des Potenzgesetzes, bestimmt in den Allelfrequenz-Spektren klassifizierter Daten. (Steigung in Doppelt-Logarithmischem Plot). Ergebnisse in TAB. 3.3.
a_z	Exponent des Potenzgesetzes, bestimmt durch lineare Regression in den kumulierten Häufigkeitsverteilungen (Zipf-Plot). Ergebnisse in TAB. 3.3.
\bar{t}_e	Durchschnittliche Anzahl von Generationen, nach dem ein Individuum die letzte Reihe des Gitters erreicht hat.
\bar{t}_f	Durchschnittliche Anzahl von Generationen, innerhalb der alle Deme des Gitters voll besiedelt waren.

nach einer Range-Expansion machen konnte (siehe dazu ABSCHN. 3.3). In den $\Gamma(p_f)$ -Spektren war ein Bereich zu identifizieren, in dem die Allelfrequenzen von SV und OM gleichermaßen durch die Range-Expansion geprägt waren. Dort habe ich den Wert „ Γ “ gemessen (siehe auch ABB. 3.4).

In Simulationen auf dem großen Gitter war in der doppelt-logarithmischen Darstellung der Allelfrequenz-Spektren eine Gerade zu erkennen. Das deutet auf ein Potenzgesetz hin, dessen Exponent durch Ermittlung der Steigung a der Geraden bestimmt werden kann. Die Steigungen habe ich durch lineare Regression sowohl in den Dichte-Verteilungen (a_k ; siehe ABB. B.11) als auch in den kumulierten Häufigkeitsverteilungen (a_z ; siehe ABB. B.12) ermittelt, um eventuelle Ungenauigkeiten durch die Klassifizierung der Daten erkennen zu können (Newman, 2005).

3. Ergebnisse

Bisherige Studien konnten zeigen, dass nicht nur positive, sondern auch neutrale und sogar nachteilige Mutationen während einer Range-Expansion hohe Frequenzen erreichen können (Edmonds et al., 2004; Klopstein et al., 2006; Travis et al., 2007). Dieses als *Gene-Surfing* bezeichnete Phänomen wurde mit Computersimulationen untersucht, in denen neue Mutationen nahe der Ausbreitungsfront eingebracht wurden. Verschiedene genetische Varianten sind aber bereits vor der Range-Expansion in der Population vorhanden (*standing genetic variation*) oder passieren während der Ausbreitung auch abseits der Front. Die Frage war daher, ob das Surfing-Phänomen in den Allelfrequenz-Spektren festgestellt werden kann und inwieweit sich der Einfluss von Standing-genetic-Variation und Ongoing-Mutations quantifizieren lässt.

Zunächst habe ich verschiedene demographische Szenarien simuliert und deren Einfluss auf die Allelfrequenz-Spektren betrachtet. Dabei zeigte sich, dass ausschließlich bei den Simulationen von Range-Expansions eine Häufung von hohen Allelfrequenzen beobachtet werden kann (ABSCHN. 3.1). Die dabei entstehenden sehr breiten Verteilungen habe ich sowohl bei Ongoing-Mutations als auch bei Standing-genetic-Variation gefunden. Darüber hinaus konnte ich zeigen, dass die Allelfrequenz-Spektren in weiten Teilen proportional zur Anfangsfrequenz p_s bzw. zur Mutationsrate μ sind (ABSCHN. 3.2).

Weitere Analysen deuten darauf hin, dass die Allele mit hohen Frequenzen p_f sehr wahrscheinlich aus Standing-genetic-Variation stammen. Demnach waren Mutationen, die durch Range-Expansions sehr häufig geworden sind, in der Regel bereits vor der Ausbreitung in der Population vorhanden (ABSCHN. 3.3).

In Simulationen auf einer Europakarte konnte ich einen starken Einfluss der geographischen Struktur auf die Allelfrequenz-Spektren von Standing-genetic-Variation beobachten. Im Gegensatz dazu veränderten sich die Allelfrequenz-Spektren in Simulationen von Ongoing-Mutations kaum (ABSCHN. 3.4).

Ergebnisse aus Simulationen mit langen Laufzeiten (Große Welt) deuten darauf hin, dass die Form der Allelfrequenz-Spektren nach einer Range-Expansion durch ein Potenzgesetz

3. Ergebnisse

(*power law*) beschrieben werden kann. Der Exponent a des Potenzgesetzes ist dabei robust gegenüber Parameteränderungen der Range-Expansion (ABSCHN. 3.5).

In ABSCHN. B.2 präsentiere ich zudem Argumente dafür, dass der theoretisch erwartete Exponent dieses Potenzgesetzes zwischen $a = -\frac{4}{3}$ und $a = -\frac{7}{5}$ liegt. Dieser Wert a liegt damit zwischen dem Exponenten einer durchmischten (*well-mixed*) exponentiell wachsenden Population ($a = -2$) und dem Exponenten einer durchmischten Population stabiler Größe ($a = -1$; Durrett (2013); Wakeley (2008)). Die Ergebnisse aus der Analyse echter genetischer Daten finden sich im Anhang in ABSCHN. B.3.

3.1. Einfluss demographischer Szenarien auf Allelfrequenz-Spektren

Für Simulationen von Standing-genetic-Variation erwartet man für lange Laufzeiten ($g \rightarrow \infty$), dass eine der genetischen Varianten fixiert wird (Hartl et al., 2007). Die Allelfrequenz-Spektren $f(p_f)$ solch langer Simulationen sind daher bimodal, mit Peaks bei $p_f = 0$ und $p_f = 1$. Da die Fixationswahrscheinlichkeit einer neutralen Mutation genau ihrer Anfangsfrequenz p_s entspricht (Hartl et al., 2007), erwartet man für die Funktionswerte der Allelfrequenz-Spektren an diesen Stellen daher $f(1) = p_s$ und $f(0) = 1 - p_s$. In dieser Arbeit habe ich untersucht, wie die Allelfrequenz-Spektren nach relevanten Simulationszeiten aussehen.

Ich habe Simulationen für drei verschiedene demographische Szenarien, jeweils für Standing-genetic-Variation und Ongoing-Mutations, durchgeführt. ABB. 3.1 und ABB. 3.2 zeigen die resultierenden Allelfrequenz-Spektren nach knapp 2000 Generationen. In den stabilen demographischen Szenarien (*Stabile Demographie* und *Räumlich stabiles Wachstum*) sind die Verteilungen nach dieser Zeit qualitativ identisch. Dabei sind die Spektren beim räumlich stabilen Wachstum nach der betrachteten Zeit etwas breiter (siehe dazu auch ABB. B.2 und ABB. B.3).

Im Gegensatz zu den Spektren der beiden stabilen demographischen Szenarien können in den Simulationen der Range-Expansions durch Gene-Surfing viel höhere Frequenzen in der betrachteten Zeit erreicht werden. So gab es Simulationen von Standing-genetic-Variation, bei denen das abgeleitete Allel nach einer der Range-Expansion in mehr als 80% der Individuen zu finden war, obwohl es zu Beginn nur eine Frequenz von $p_s = 1\%$ hatte (siehe ABB. 3.1). In den Simulationen der beiden anderen Szenarien war die maximale

3.2. Einfluss von Mutationsrate und Anfangsfrequenz auf Allelfrequenz-Spektren

Frequenz hingegen 9.2% (*Standing-genetic-Variation*) bzw. 5.6% (*Ongoing-Mutations*). Daher sind in den Spektren der gezeigten Simulationen in Bereichen hoher Allelfrequenzen (> 10%) ausschließlich Surfer aus Range-Expansions zu finden.

Die hohen Frequenzen lassen sich durch die Sektoren erklären, die sich nur bei Range-Expansions, aber nicht bei einer stabilen Demographie oder dem räumlich stabilen Wachstum bilden (siehe auch ABB. A.2). Hat sich ein solcher Sektor an der Front ausgebildet, so erfolgen weitere Besiedlungen ausschließlich von Trägern einer einzigen genetischen Variante. Messbar ist dies auch durch einen starken Anstieg des Inzucht-Koeffizienten (F_{st} -Wert; Holsinger and Weir (2009)). Dieser Anstieg ist bei Range-Expansions deutlich stärker als bei den beiden anderen demographischen Szenarien.

3.2. Einfluss von Mutationsrate und Anfangsfrequenz auf Allelfrequenz-Spektren

Bei Standing-genetic-Variation erwartet man, dass verschiedene Genorte unterschiedliche Anfangsfrequenzen p_s für das abgeleitete Allel haben können. So kann z.B. an einem Genort das abgeleitete Allel in jedem hundertsten Individuum der Population zu finden sein ($p_s = 10^{-2}$), eine erst kürzlich aufgetretene Mutation findet sich hingegen nur in jedem zehntausendsten Individuum ($p_s = 10^{-4}$). Daher sollten bei einer höheren Mutationsrate oder einer hohen Anfangsfrequenz mehr Mutationen an der Front auftreten bzw. vorhanden sein. Somit sollte es zu mehr Surfing-Ereignissen kommen.

Um die genauen Auswirkungen von μ und p_s auf die Allelfrequenz-Spektren zu untersuchen, habe ich Range-Expansions für verschiedene Mutationsraten und Anfangsfrequenzen simuliert. Dabei führte eine zehnfache Mutationsrate bzw. Anfangsfrequenz zu im Mittel zehnmal so vielen Surfing-Ereignissen (siehe ABB. B.4). Teilt man nun die Allelfrequenz-Spektren durch die jeweils verwendete Mutationsrate μ bzw. die Anfangsfrequenz p_s (Normalisierung), so lassen sich die Verteilungen zur Deckung bringen. Wie in ABB. 3.3 zu sehen ist funktioniert das insbesondere für die vom Gene-Surfing betroffenen hohen Frequenzen p_f sehr gut. Aufgrund dieser direkten Proportionalität können (verallgemeinerte) Funktionen γ_{OM} und γ_{SV} postuliert werden, aus denen man die Allelfrequenz-Spektren (f_{OM} und f_{SV}) für jede Mutationsrate μ bzw. Anfangsfrequenz p_s ableiten kann (\hat{S} : weitere Parameter der Simulation wie z.B. die Migrationsrate m oder die Kapazitätsgrenze K).

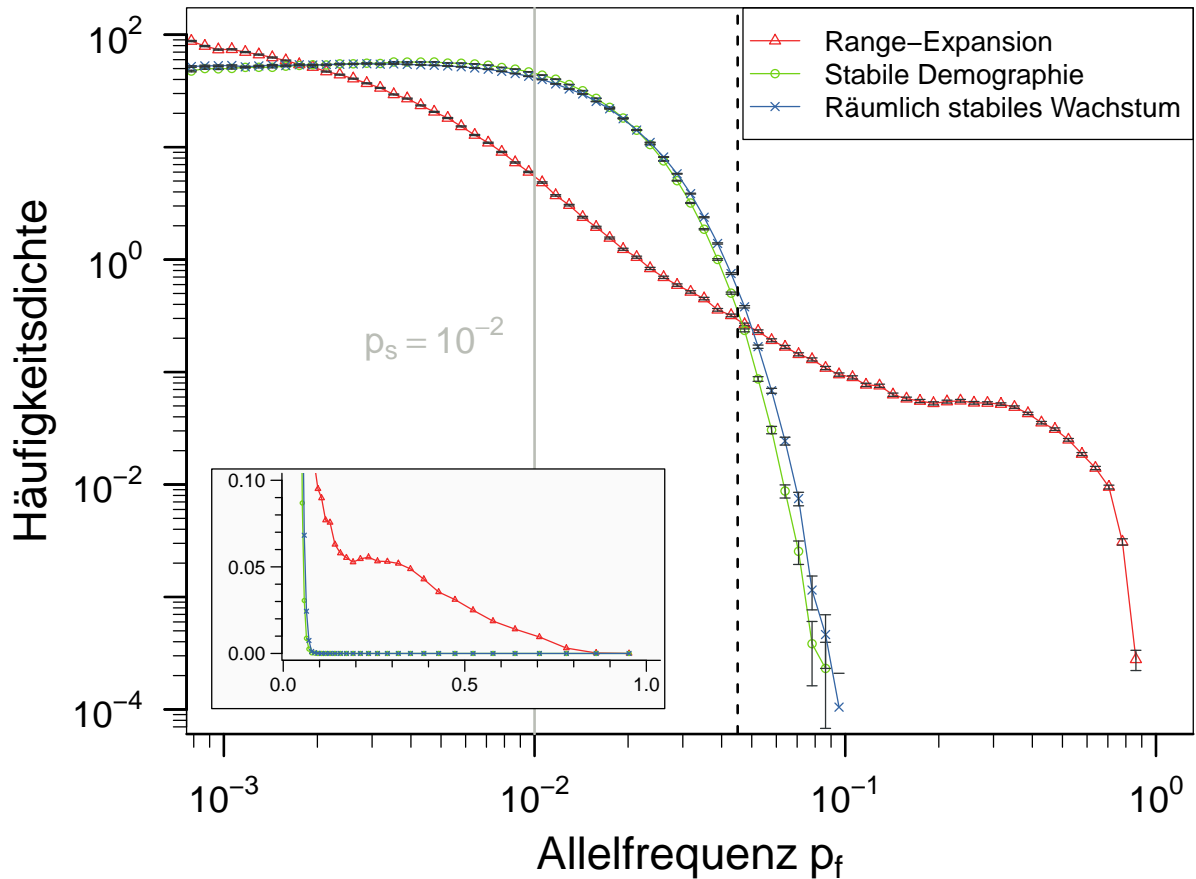


Abbildung 3.1.: **Auswirkungen von drei demographischen Szenarien auf die Allelfrequenz-Spektren von Standing-genetic-Variation.** Im Laufe einer Range-Expansion (rote Dreiecke) können anfänglich seltene genetische Varianten ($p_s = 10^{-2}$; vertikale graue Linie) sehr hohe Frequenzen ($p_f^{max} \approx 0.87$) im neu besiedelten Bereich erreichen. Innerhalb dieses relativ kurzen Zeitraumes findet also eine massive Häufung hoher Allelfrequenzen statt. Im Gegensatz dazu tritt im selben Zeitraum ($g = 1990$ Generationen) bei einer stabilen Demographie (grüne Kreise) und beim räumlich stabilen Wachstum (blaue Kreuze) keine wesentliche Häufung hoher Allelfrequenzen auf (hier: $p_f^{max} \approx 0.09$). Auch die Anzahl $n_{>}$ der Simulationen, in denen eine hohe Frequenz (hier: $p_f > 0.045$; gestrichelte Linie) erreicht wurden, ist bei Range-Expansions deutlich höher ($n_{>} = 31440$) als bei Simulationen der stabilen Demographie ($n_{>} = 711$) und dem räumlich stabilen Wachstum ($n_{>} = 1499$). Die Fehlerbalken geben den Standardfehler beim Bootstrapping der Verteilung an (siehe TEILABSCHN. A.3.1.2). Die eingefügte Grafik zeigt das gleiche Allelfrequenz-Spektrum mit linear-skalierten Achsen. Die demographischen Szenarien sind in TEILABSCHN. 2.1.4 genauer beschrieben. Für jedes der drei gezeigten Allelfrequenz-Spektren wurden 10^6 Simulationen auf der Kunstwelt durchgeführt. Verwendete Parameter: $K = 100$, $m = 0.2$, $r = 0.5$. Beim räumlich stabilen Wachstum: $K_{start} = 10.19$, $r_{deme} = 0.235$. Ein Vergleich der Allelfrequenz-Spektren nach $g = 400$ Generationen befindet sich im Anhang (siehe ABB. B.2).

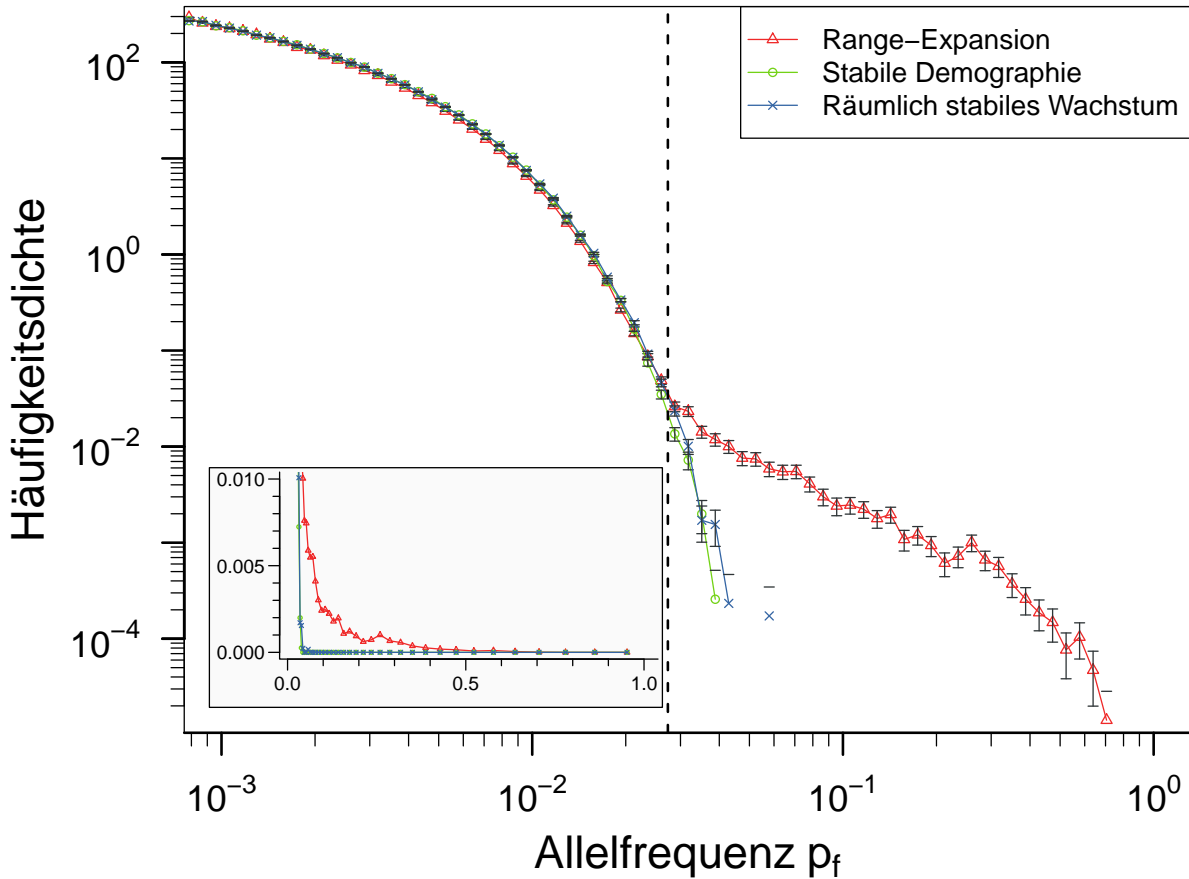


Abbildung 3.2.: **Auswirkungen von drei demographischen Szenarien auf die Allelfrequenz-Spektren von Ongoing-Mutations.** Während einer Range-Expansion (rote Dreiecke) können neu auftretende Mutationen (Mutationsrate $\mu = 10^{-6}$) sehr hohe Frequenzen ($p_f^{max} \approx 0.74$) im neu besiedelten Bereich erreichen. Im Vergleich dazu sind in gleich langen Simulationen ($g = 1990$ Generationen) einer stabilen Demographie (grüne Kreise) und räumlich stabilem Wachstum (blaue Kreuze) die maximalen Allelfrequenzen am Ende der Simulation deutlich geringer ($p_f^{max} \approx 0.038$ bzw. $p_f^{max} \approx 0.056$). Siehe dazu auch ABB. 3.1. Auch die Anzahl $n_{>}$ der Simulationen, in denen eine hohe Frequenz (hier: $p_f > 0.027$; gestrichelte Linie) erreicht wurden, ist bei Range-Expansions deutlich höher ($n_{>} = 781$) als bei Simulationen der stabilen Demographie ($n_{>} = 31$) und dem räumlich stabilen Wachstum ($n_{>} = 46$). Die Übereinstimmung zwischen den Szenarien links der senkrechten gestrichelten Linie entspricht unserer Erwartung für Mutationen abseits der Front. Wie in ABB. 3.1 geben die Fehlerbalken den Standardfehler beim Bootstrapping der Verteilung an (siehe TEILABSCHN. A.3.1.2). Die eingefügte Grafik zeigt das gleiche Allelfrequenz-Spektrum mit linear-skalierten Achsen. Die demographischen Szenarien sind in TEILABSCHN. 2.1.4 genauer beschrieben. Für jedes der drei gezeigten Allelfrequenz-Spektren wurden 10^6 Simulationen auf der Kunstwelt durchgeführt. Verwendete Parameter: $K = 100$, $m = 0.2$, $r = 0.5$. Beim räumlich stabilen Wachstum: $K_{start} = 10.19$, $r_{deme} = 0.235$. Ein Vergleich der Allelfrequenz-Spektren nach $g = 400$ Generationen befindet sich im Anhang (siehe ABB. B.3).

$$f_{OM} \left(p_f | \mu, \hat{S} \right) = \mu \cdot \gamma_{OM} \left(p_f | \hat{S} \right) \quad (3.1)$$

$$f_{SV} \left(p_f | p_s, \hat{S} \right) = p_s \cdot \gamma_{SV} \left(p_f | \hat{S} \right) \quad (3.2)$$

Bei hohen Mutationsraten oder Anfangsfrequenzen kann eine Deckung im Bereich niedriger Allelfrequenzen hingegen nicht erreicht werden (ABB. 3.3), da niedrige Allelfrequenzen mit steigender Mutationsrate bzw. bereits sehr hoher Anfangsfrequenz weniger wahrscheinlich sind.

3.3. Beitrag von Standing-genetic-Variation und Ongoing-Mutations zu den Allelfrequenz-Spektren

Wie bereits in der Einleitung (ABSCHN. 1.4) erwähnt treten in tatsächlichen Populationen Standing-genetic-Variation und Ongoing-Mutations gleichzeitig auf. Die Frage ist nun, mit welcher Wahrscheinlichkeit die gesurften Allele im Bereich hoher Allelfrequenzen aus der Ursprungspopulation (Standing-genetic-Variation) stammen bzw. während der Range-Expansion aufgetreten sind. Wie im ABSCHN. 3.2 beschrieben hängt die relative Wichtigkeit dabei von der Mutationsrate μ bzw. der Anfangsfrequenz p_s ab.

Angenommen, wir kennen μ und p_s , so kann man die Frage stellen: Wie groß ist die Wahrscheinlichkeit $P_{ancestral}$, dass eine Mutation mit der Frequenz p_f aus Standing-genetic-Variation stammt? Die Wahrscheinlichkeit entspricht genau dem Anteil, den das Spektrum von SV (f_{SV}) an einem Spektrum ($f_{SV} + f_{OM}$) hat, bei dem sowohl SV als auch OM auftreten:

$$\begin{aligned} P_{ancestral} \left(p_f | \mu, p_s, \hat{S} \right) &= \frac{f_{SV} \left(p_f | p_s, \hat{S} \right)}{f_{SV} \left(p_f | p_s, \hat{S} \right) + f_{OM} \left(p_f | \mu, \hat{S} \right)} \\ &= \frac{1}{1 + \frac{f_{OM} \left(p_f | \mu, \hat{S} \right)}{f_{SV} \left(p_f | p_s, \hat{S} \right)}}. \end{aligned} \quad (3.3)$$

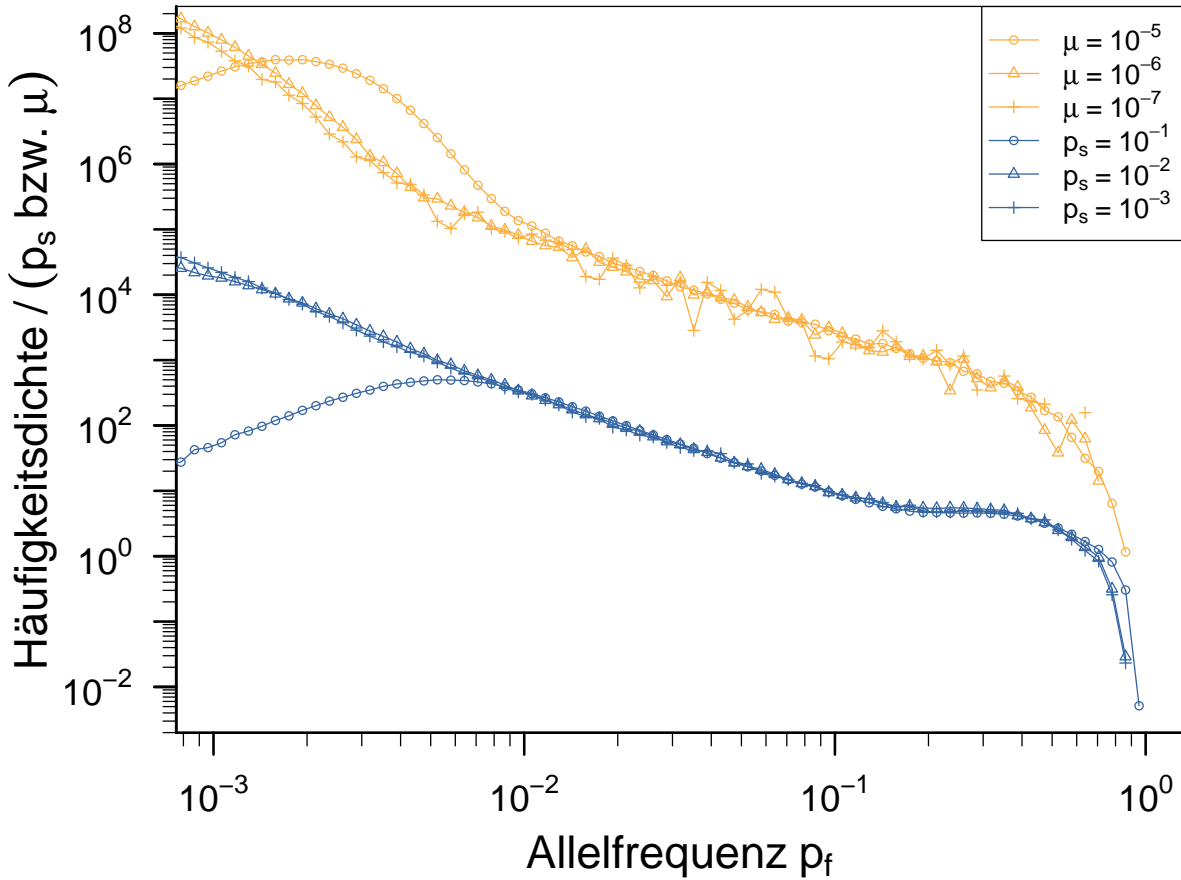


Abbildung 3.3.: **Normalisierte Allelfrequenz-Spektren aus Range-Expansion Simulationen für Standing-genetic-Variation und Ongoing-Mutations.** Die Allelfrequenz-Spektren können zur Deckung gebracht werden, wenn man die Häufigkeitsdichten durch die verwendete Anfangsfrequenz p_s bzw. die Mutationsrate μ teilt (*Normalisierung*). Diese Überlagerung funktioniert besonders gut in den Bereichen hoher Allelfrequenzen p_f . Der überlappende Bereich entspricht den universellen Funktionen γ_{OM} (GL. 3.1) bzw. γ_{SV} (GL. 3.2). Mit steigenden Anfangsfrequenzen p_s bzw. hohen Mutationsraten μ werden niedrige Allelfrequenzen p_f unwahrscheinlicher, so dass eine Überlagerung insbesondere in den Bereichen niedriger Allelfrequenzen p_s nicht möglich ist. Die Allelfrequenzen für die gezeigten Spektren wurden nach $g = 400$ Generationen gemessen, also direkt nach der Range-Expansion (siehe TAB. 3.2). Die nicht zur Deckung gebrachten Spektren sind mit ihren Standardfehlern in ABB. B.4 gezeigt. Für jedes der sechs gezeigten Allelfrequenz-Spektren wurden 10^6 Simulationen auf der Kunstwelt durchgeführt.

3. Ergebnisse

Wie in [ABB. 3.3](#) gezeigt können verallgemeinerte Allelfrequenz-Spektren formuliert werden, die unabhängig von μ und p_s sind (siehe [GL. 3.1](#) und [GL. 3.2](#)). Setzt man diese Formeln in [GL. 3.3](#) ein, so wird ein entscheidender Teil der Gleichung (im Folgenden $\Gamma(p_f)$ genannt) unabhängig von μ bzw. p_s :

$$\begin{aligned} P_{ancestral}(p_f|\mu, p_s, \hat{S}) &= \frac{1}{1 + \frac{\mu}{p_s} \cdot \frac{\gamma_{OM}(p_f|\hat{S})}{\gamma_{SV}(p_f|\hat{S})}} \\ &= \frac{1}{1 + \underbrace{\frac{\mu}{p_s} \cdot \Gamma(p_f)}_{\Lambda}}. \end{aligned} \quad (3.4)$$

Der Ausdruck $\Gamma(p_f)$ entspricht genau dem Verhältnis der verallgemeinerten Formeln aus [GL. 3.1](#) und [GL. 3.2](#). Daher können wir die Funktion $\Gamma(p_f)$ aus den Simulationsdaten ermitteln, in dem wir ein normalisiertes Spektrum von Ongoing-Mutations durch ein normalisiertes Spektrum von Standing-genetic-Variation dividieren (siehe [ABB. 3.4](#)). Mit Hilfe des experimentell bestimmten $\Gamma(p_f)$ -Wertes können wir nun Aussagen über den relativen Beitrag von Ongoing-Mutations und Standing-genetic-Variation zum Allelfrequenz-Spektrum machen.

Nimmt $P_{ancestral}(p_f)$ den Wert $\frac{1}{2}$ an, so ist es gleich wahrscheinlich, dass eine Mutation mit einer Frequenz p_f aus SV oder OM stammt. Das ist genau dann der Fall, wenn $\frac{p_s}{\mu} = \Gamma(p_f)$ und somit $\Lambda = 1$ ist. Wird allerdings $\Gamma(p_f) \cdot \mu > p_s$ und damit $\Lambda > 1$ (z.B. aufgrund eines hohen Wertes für $\Gamma(p_f)$), so wird $P_{ancestral} < \frac{1}{2}$. In diesem Falle sind Allele mit der Frequenz p_f mit einer größeren Wahrscheinlichkeit während der Range-Expansion entstanden (siehe auch [TAB. 3.1](#)). Ein steigender $\Gamma(p_f)$ -Wert deutet somit auf eine größere Bedeutung von Ongoing-Mutations hin.

In [ABB. 3.4](#) erkennt man, dass $\Gamma(p_f)$ im Prinzip von der Allelfrequenz p_f abhängt. So spielen für kleine Allelfrequenzen p_f Ongoing-Mutations eine größere Rolle, während in den Bereichen nahe $p_f = 1$ vor allem gesurft Allele aus der Ursprungspopulation zu finden sind. Allerdings gibt es einen Bereich (*Abschnitt (II)* in [ABB. 3.4](#)), in dem $\Gamma(p_f)$ nahezu unabhängig von der Allelfrequenz p_f ist. Den Funktionswert dieses Bereiches bezeichne ich als Γ . In diesem Bereich werden Allele aus der Ursprungspopulation und solche, die während der Ausbreitung entstehen, gleichermaßen von der Range-Expansion beeinflusst. Dadurch haben die Allelfrequenz-Spektren für SV und OM in diesem Abschnitt gleiche Steigung, die ein spezifisches Merkmal von Range-Expansions ist (siehe auch

3.3. Beitrag von Standing-genetic-Variation und Ongoing-Mutations zu den Allelfrequenz-Spektren

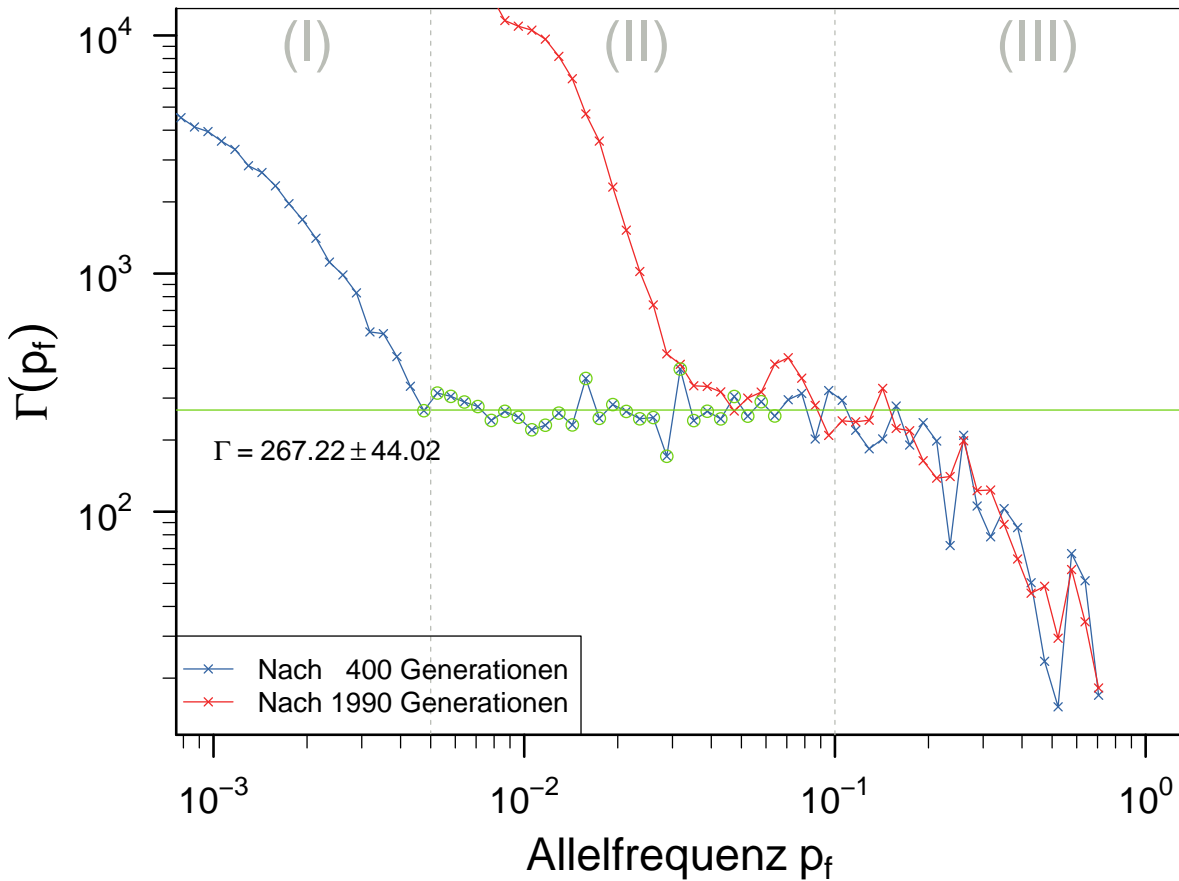


Abbildung 3.4.: **Bedeutung von Ongoing-Mutations für das Allelfrequenz-Spektrum während und nach einer Range-Expansion.** Gezeigt ist der Quotient $\Gamma(p_f)$ aus den normalisierten Allelfrequenz-Spektren von Standing-genetic-Variation und Ongoing-Mutations. Ein großer Wert für $\Gamma(p_f)$ deutet auf einen starken Einfluss von Ongoing-Mutations für die entsprechenden Allelfrequenzen p_f hin. Im *Abschnitt (I)* finden sich Allele mit niedrigen Frequenzen. Diese können im Verlauf der Simulation bereits durch Mutation einzelner Individuen auch abseits der Front erreicht werden. *Abschnitt (II)* zeigt den Bereich der Allelfrequenz-Spektren, in dem die Frequenzen der Allele aus Standing-genetic-Variation und Ongoing-Mutations gleichermaßen von der Range-Expansion betroffen sind. Ein Wert von $\Gamma = 267$ bedeutet, dass in diesem Frequenzbereich hauptsächlich Mutationen aus der Ursprungspopulation zu finden sind, sofern diese eine Frequenz von $p_s > 2.67 \times 10^{-6}$ hatten (angenommene Mutationsrate: $\mu = 10^{-8}$). *Abschnitt (III)*: Allele mit Frequenzen von $p_f > 10^{-1}$ waren mit sehr großer Wahrscheinlichkeit bereits in der Ausgangspopulation vorhanden. Die Zunahme des Anteils von niedrigfrequenten Allelen aus Ongoing-Mutations mit der Zeit (rote Kurve) ist darauf zurückzuführen, dass durch mehrfache Mutation des gleichen Genortes sowie durch genetische Drift höhere Frequenzen erreicht werden können. Im Bereich hoher Frequenzen (hier: $p_f > 3 \times 10^{-2}$) gibt es mit der Zeit jedoch kaum eine Veränderung. $\Gamma(p_f)$ wurde aus Simulationen für $\mu = 10^{-6}$ und $p_s = 10^{-3}$ direkt im Anschluss an die Range-Expansion ($g = 400$ Generationen) sowie nach $g = 1990$ Generationen berechnet. Grün eingekreiste Punkte wurden für die Berechnung des Mittelwertes verwendet. ($K = 100$; $m = 0.2$, $r = 0.5$; Gitter: *Kunstwelt*)

3. Ergebnisse

Tabelle 3.1.: Aussage des Gamma-Wertes über die relative Bedeutung von Ongoing-Mutations und Standing-genetic-Variation.

$\Gamma(p_f)$	Λ	$P_{ancestral}(p_f)$	Bedeutung
$> \frac{p_s}{\mu}$	> 1	$< \frac{1}{2}$	Für diese Werte von μ und p_s stammen die Allele mit der Frequenz p_f mit einer größeren Wahrscheinlichkeit aus Ongoing-Mutations.
$= \frac{p_s}{\mu}$	$= 1$	$= \frac{1}{2}$	Bei diesem Wert für $\Gamma(p_f)$ haben 50% der Allele mit der Frequenz p_f ihren Ursprung in der Ausgangspopulation. Die andere Hälfte der Allele ist während der Range-Expansion entstanden.
$< \frac{p_s}{\mu}$	< 1	$> \frac{1}{2}$	Es ist wahrscheinlicher, dass die Allele mit der Frequenz p_f bereits in der Ursprungspopulation vorhanden waren als dass sie während der Range-Expansion entstanden sind.

ABSCHN. 3.5). Insbesondere in Simulationen mit einer niedrigen Mutationsrate μ ist mit der Zeit kaum eine Änderung der Spektren im Bereich hoher Allelfrequenzen p_f zu beobachten.

Um eine Aussage über die relative Wichtigkeit von SV und OM für die Allelfrequenz-Spektren treffen zu können, habe ich die Γ -Werte des Plateaus (siehe ABB. 3.4) für verschiedene Parameter (Wachstumsrate, Migrationsrate etc.) gemessen. Alle Werte lagen in der Größenordnung $10^2 - 10^3$ (siehe TAB. 3.2). Dabei waren die Werte für Γ weniger von den jeweiligen Parametern der Simulation abhängig als vielmehr von der Dauer \bar{t}_f der Range-Expansion (siehe dazu ABB. B.5). Nun kann man diejenige Anfangsfrequenz p_s suchen, bei der OM und SV gleich wichtig sind ($\Lambda = 1$). Bei einer angenommenen Mutationsrate von $\mu = 10^{-8}$ (1000 Genomes Project Consortium, 2010; Conrad et al., 2011) tragen also all diejenigen Allele, die in der Ausgangspopulation eine Frequenz von $p_s > 10^{-5}$ hatten, mehr zum Allelfrequenz-Spektrum bei als Mutationen, die während der Range-Expansion entstanden sind (vergleiche TAB. 3.1).

3.4. Einfluss der geographischen Struktur auf Allelfrequenz-Spektren

Die bisher gezeigten Ergebnisse stammten aus Simulationen auf der *Kunstwelt* (siehe ABB. A.2 und TAB. 2.2), einem homogenen rechteckigen Habitat, wie es bereits in

Tabelle 3.2.: **Laufzeiten und Γ -Werte von Simulationen auf der Kunstwelt.** Gezeigt ist die durchschnittliche Anzahl von Generationen (\bar{t}_e), nach denen ein Individuum die letzte Reihe des Gitters erreicht hat bzw. nach denen alle Deme bis zur Kapazitätsgrenze besiedelt waren (\bar{t}_f). Die Γ -Werte wurden zum Zeitpunkt $\lceil \frac{\bar{t}_f}{10} \rceil \times 10$ im Bereich des Plateaus der $\Gamma(p_f)$ -Funktion gemessen (siehe ABB. 3.4). Die dafür verwendeten Allelfrequenz-Spektren wurden aus je 2×10^6 Simulationen ($\mu = 10^{-6}$ bzw. $p_s = 10^{-3}$) auf der Kunstwelt durchgeführt. Die Werte für Γ sind linear abhängig von der Laufzeit der Simulationen (siehe ABB. B.5). In der Spalte *Parameter* sind diejenigen Parameter aufgeführt, deren Werte sich von den Standardwerten ($K = 100$, $m = 0.2$, $r = 0.5$, $\alpha = n.a.$, $\epsilon = n.a.$; siehe auch TAB. 2.3) unterscheiden. (*SD* = Standardabweichung). Die Laufzeiten für Simulationen auf der Europakarte sind in TAB. B.1 aufgeführt.

Parameter	\bar{t}_e	<i>SD</i> [t_e]	\bar{t}_f	<i>SD</i> [t_f]	Γ	<i>SD</i> [Γ]
Standard Parameter	336.2	5.99	368.9	5.28	260.2	28.7
$K = 20$	379.6	8.95	413.7	8.27	316.2	39.5
$K = 200$	326.5	5.24	359.7	4.54	236.6	22.6
$K = 1000$	312.8	4,08	348.3	3.44	209.5	16.2
$m = 0.04$	656.3	11.82	707.2	10.38	536.1	55.3
$m = 0.40$	247.8	4.40	276.2	3.99	177.6	18.2
$m = 0.80$	180.7	3.20	206.5	3.09	117.8	11.7
$m = 1.00$	163.6	2.88	189.2	2.73	105.3	9.9
$r = 0.25$	470.7	9.93	530.2	9.46	^a 417.2	42.8
$r = 0.75$	278.7	4.52	301.8	3.83	^a 182.1	28.5
$r = 1.00$	245.4	3.72	263.4	3.09	^a 162.3	24.8
$\alpha = 0.05, \epsilon = 2$	449.4	4.94	482.4	4.28	^a 396.0	34.3
$\alpha = 0.10, \epsilon = 10$	526.2	4.73	560.6	4.07	485.1	39.0
$\alpha = 0.20, \epsilon = 10$	679.2	6.20	721.9	4.93	662.8	45.9
$\alpha = 0.40, \epsilon = 10$	1040.5	11.96	1113.0	8.39	1160.2	67.2

^a Verwendete Allelfrequenz-Spektren wurden nur aus 1×10^6 Simulationen erstellt.

3. Ergebnisse

anderen Simulationsstudien verwendet wurde (z.B.: [Edmonds et al., 2004](#); [Klopfstein et al., 2006](#); [Travis et al., 2007](#)). In der tatsächlichen Umwelt gibt es allerdings Barrieren, die zu geographischen Engstellen (*geographic bottlenecks*) und zu ungleichmäßigen Frontverläufen führen können. Um die Gültigkeit der bisher gezeigten Ergebnisse in natürlichen Systemen abschätzen zu können, habe ich zusätzlich Simulationen auf einer *Europakarte* durchgeführt (vgl. [ABB. A.3](#)).

In [ABB. 3.5](#) vergleiche ich die der Simulationen auf der Kunstwelt mit denen auf der Europakarte, jeweils für Ongoing-Mutations und Standing-genetic-Variation. Dabei fällt im Spektrum von Standing-genetic-Variation auf, dass einige mittlere Frequenzen p_f auf der Europakarte seltener zu finden sind als in vergleichbaren Simulationen auf der Kunstwelt. Zudem gibt es einige hohe Frequenzen ($p_f > 2 \times 10^{-1}$), deren Häufigkeit auf der Europakarte im Vergleich zu den Ergebnissen auf der Kunstwelt gestiegen ist. In den Simulationen von Ongoing-Mutations ist dieser Einfluss der geographischen Struktur im Vergleich der beiden Karten hingegen nicht zu erkennen. Das deckt sich mit den Erwartungen, die man aus [Burton and Travis \(2008\)](#) ableiten kann. Entweder schafft es eine vorhandene genetische Variante durch eine Engstelle und kann durch Gene-Surfing hohe Frequenzen erreichen oder sie verbleibt hinter der Engstelle bei einer geringen Frequenz.

Die Tatsache, dass die geographische Struktur vor allem Auswirkungen auf Standing-genetic-Variation hat, zeigt sich auch in der Betrachtung des relativen Beitrags beider Mutationstypen zu den Allelfrequenz-Spektren (siehe [ABB. 3.6](#)). Der relative Anteil von Ongoing-Mutations an den Frequenzen, die in Simulationen von Standing-genetic-Variation auf der Europakarte nicht erreicht werden, steigt. Darüber hinaus ist zu erkennen, dass Ongoing-Mutations auf der simulierten Europakarte im Vergleich zur Kunstwelt mehr zum Allelfrequenz-Spektrum beitragen.

Allerdings kann auch das Verhältnis von Frontbreite und Lauflänge als „Engstelle“ interpretiert werden und führt zu ähnlichen Ergebnissen in den Allelfrequenz-Spektren von Standing-genetic-Variation. In [ABB. B.10](#) zeige ich Allelfrequenz-Spektren aus Simulationen auf einem quadratischen und auf einem langgezogenen Gitter. Dabei ist zu erkennen, dass auf dem langen Gitter ein großer Bereich von hohen Frequenzen nicht im Spektrum repräsentiert ist. Es ist offensichtlich, dass nach einer typischen Laufzeit eine Mutation entweder die Front übernommen oder den Kontakt zur Front verloren hat. Im weiteren Verlauf der Simulationen führt dies dazu, dass der relative Anteil der Mutation an der Gesamtpopulation entweder weiter steigt oder fällt. Bei Ongoing-Mutations ist

der Zeitpunkt einer Mutation hingegen nicht auf den Beginn der Simulation festgelegt. Daher ist dieser Effekt hier nicht zu erwarten.

3.5. Potenzgesetz in Simulationen auf der Großen Welt

Nach einer Range-Expansion zeigt sich in den Allelfrequenz-Spektren sowohl von Ongoing-Mutations als auch von Standing-genetic-Variation ein deutlicher Anstieg von Allelen mit hohen Frequenzen (ABSCHN. 3.1). In einem Bereich der Spektren werden neue Mutationen durch die Range-Expansion im gleichen Maße verändert wie die bereits existierenden Mutationen. Die gleiche Form der Spektren in diesem Bereich ist insbesondere in den Graphen für Γ zu sehen und scheint spezifisch für Range-Expansions zu sein (siehe ABSCHN. 3.3). Der Effekt wird vor allem bei langen Range-Expansions sichtbar, da Surfing-Ereignisse abgeschlossen werden können. Um die asymptotische Form der Spektren besser beschreiben zu können, habe ich daher Simulationen mit längeren Laufzeiten auf einem großen Gitter ($w = 1000$, $h = 1000$) durchgeführt.

In den Simulationen von Range-Expansions finde ich im doppelt-logarithmischen Plot der resultierenden Allelfrequenz-Spektren eine Gerade (siehe z.B. ABB. 3.7). Das deutet auf ein Potenzgesetz (*power law*) hin, bei dem die Häufigkeitsdichte $f(p_f)$ einer Frequenz p_f folgendermaßen abnimmt:

$$f(p_f) = C \cdot p_f^a. \quad (3.5)$$

Wie in ABSCHN. 3.2 gezeigt ist der Faktor C unter anderem linear von der Mutationsrate μ bzw. der Anfangsfrequenz p_s abhängig. Den Exponenten a dieses Potenzgesetzes habe ich für verschiedene Migrationsraten m (ABB. 3.7), Kapazitätsgrenzen K (ABB. B.9), Wachstumsraten r (ABB. B.8) und Allee-Effekte (ABB. 3.8) im doppelt-logarithmischen Plot durch lineare Regression bestimmt. In den Wahrscheinlichkeitsdichte-Verteilungen der klassifizierten Daten habe ich Exponenten zwischen $a_k = -1.32$ und $a_k = -1.46$ ermittelt (siehe TAB. 3.3). Um die Ergebnisse zu verifizieren, habe ich die Steigungen zusätzlich in Zipf-Plots gemessen und dort Steigungen zwischen $a_z = -0.35$ und $a_z = -0.45$ ermittelt (siehe TAB. 3.3). Dabei gilt: $a_k = a_z - 1$ (siehe TEILABSCHN. A.3.1.4 und Newman (2005)).

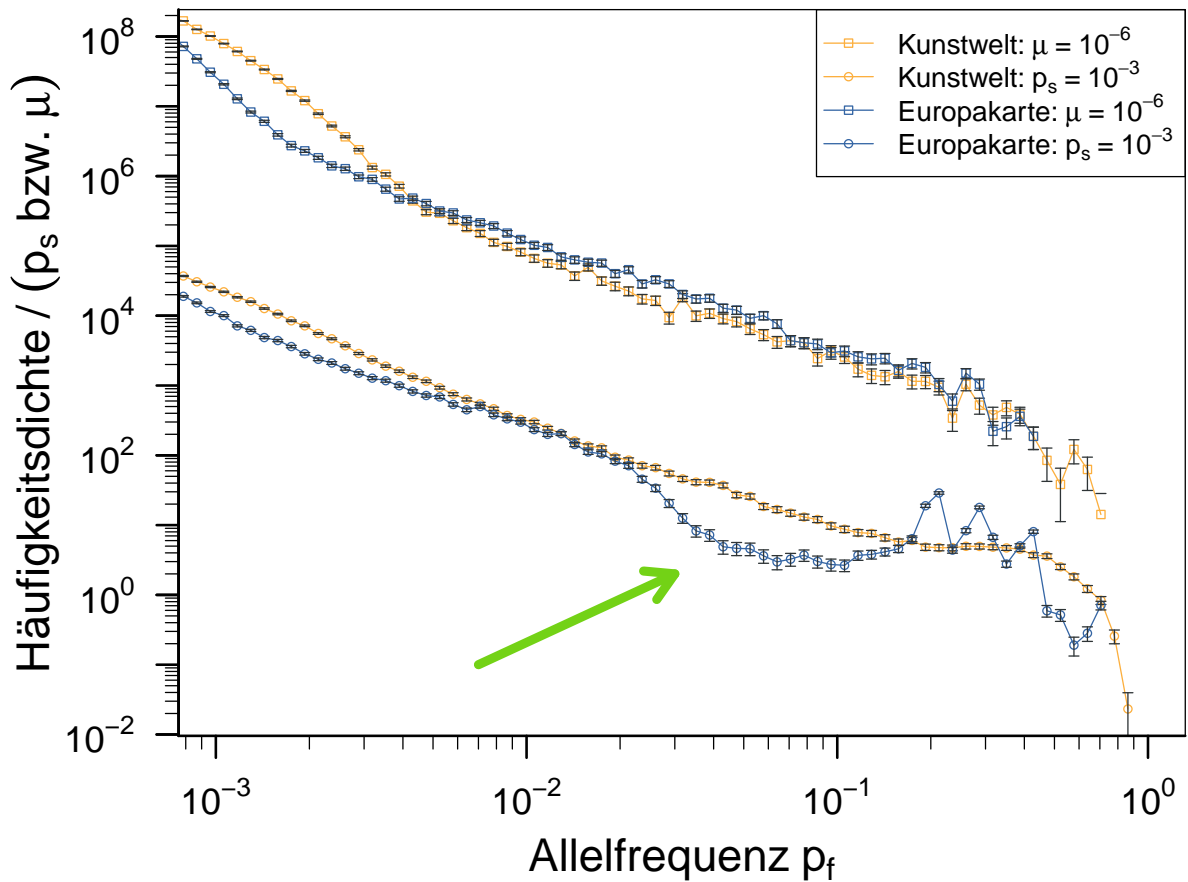


Abbildung 3.5.: **Normalisierte Allelfrequenz-Spektren aus Range-Expansion Simulationen auf der Kunstwelt und der Europakarte.** Im Gegensatz zu Ongoing-Mutations (Vierecke) ist in den Allelfrequenz-Spektren von Standing-genetic-Variation (Kreise) ein deutlicher Einfluss der geographischen Struktur zu erkennen (grüner Pfeil). Aufgrund geographischer Engstellen (*geographic bottlenecks*) werden einige Frequenzen sehr unwahrscheinlich. Erreicht die Population z.B. den Bosphorus, ist es sehr wahrscheinlich, dass lediglich eine der vorhandenen genetischen Varianten nach der Passage noch in der Population vorhanden ist. Befindet sich die Engstelle am Anfang der Range-Expansion, so erreichen die genetischen Varianten aus der Ursprungspopulation in der Regel entweder sehr hohe oder sehr niedrige Frequenzen. Bei Ongoing-Mutations kann die jeweils andere genetische Variante hingegen auch nach dem Verlust erneut in der Population auftreten und surfen. Die Allelfrequenzen für die gezeigten Spektren wurden aus 10^6 Simulationen auf der Kunstwelt nach $g = 400$ Generationen gemessen, also direkt nach der Range-Expansion (siehe TAB. 3.2 und TAB. B.1). Die Fehlerbalken geben den Standardfehler beim Bootstrapping der Verteilung an (siehe TEILABSCHN. A.3.1.2). Normalisierte Spektren für verschiedene μ und p_s von Simulationen auf der Europakarte sind in ABB. B.6 dargestellt.

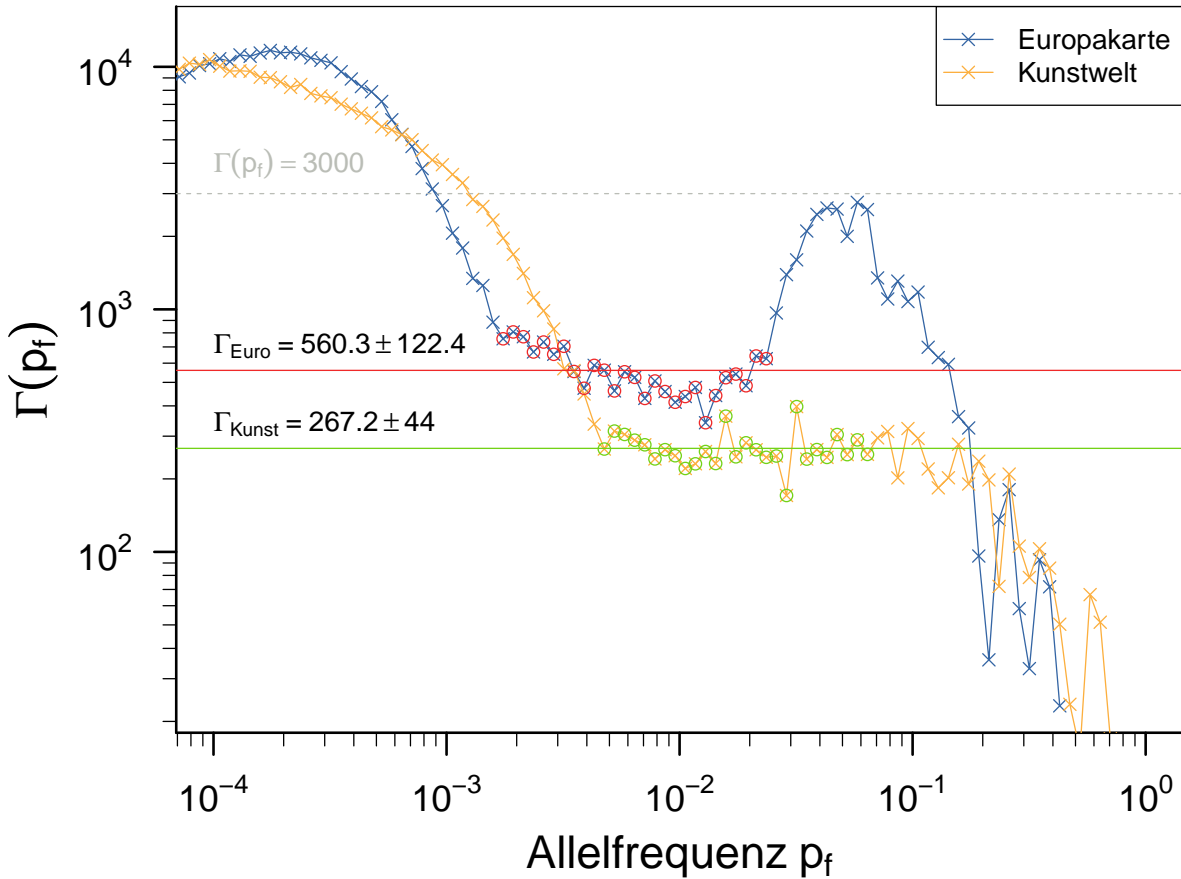


Abbildung 3.6.: **Einfluss der geographischen Struktur auf den relativen Anteil von Ongoing-Mutations am Allelfrequenz-Spektrum.** Gezeigt sind die Graphen für $\Gamma(p_f)$ aus Simulationen auf der Kunstwelt und auf der Europakarte. Je höher der Wert für $\Gamma(p_f)$, desto größer ist der Anteil von neuen Mutationen am Allelfrequenz-Spektrum. Generell ist der Anteil von Ongoing-Mutations am Allelfrequenz-Spektrum auf der Europakarte höher als auf der Kunstwelt. Einige Frequenzen p_f werden in Simulationen von Standing-genetic-Variation selten erreicht. Dadurch steigt in diesem Bereich der relative Einfluss von Ongoing-Mutations (Scheitelpunkt bei $p_f = 4 \times 10^{-2}$). Allelfrequenzen wurden nach $g = 400$ Generationen gemessen. ($K = 100$, $m = 0.2$, $r = 0.5$)

3. Ergebnisse

Tabelle 3.3.: Exponenten des Potenzgesetzes in Allelfrequenz-Spektren von Simulationen mit Standing-genetic-Variation. Die Simulationen wurden so lange auf der *Großen Welt* durchgeführt, bis ein Individuum die letzte Reihe erreicht hat (t_e). Die roten Werte sind diejenigen Parameter, die sich von den Standardwerten ($m = 0.2$, $K = 100$, $r = 0.5$, $\alpha = n.a.$, $\epsilon = n.a.$; siehe auch TAB. 2.3) unterscheiden. Gemessen wurde die Steigung in Allelfrequenz-Spektren klassifizierter Daten (a_k) bzw. Zipf-Plots (a_z). Die Bestimmtheitsmaße (R^2) der linearen Regression lagen bei beiden Methoden nahe Eins (klassifizierte Daten: $R^2 > 0.997$; Zipf-Plots: $R^2 > 0.999$). Beispiele für die Messung der Exponenten sind in ABB. B.11 und ABB. B.12 dargestellt.

K	m	r	α	ϵ	\bar{t}_e	$SD[t_e]$	a_k	a_z
20	0.2	0.5	<i>n.a.</i>	<i>n.a.</i>	3839.1	8.2	-1.344	-0.356
20	0.4	0.5	<i>n.a.</i>	<i>n.a.</i>	2742.9	5.3	-1.356	-0.367
20	0.8	0.5	<i>n.a.</i>	<i>n.a.</i>	1965.5	3.5	-1.407	-0.393
100	0.2	0.5	<i>n.a.</i>	<i>n.a.</i>	3431.4	6.8	-1.346	-0.362
100	0.4	0.5	<i>n.a.</i>	<i>n.a.</i>	2537.5	4.9	-1.380	-0.384
100	0.8	0.5	<i>n.a.</i>	<i>n.a.</i>	1860.3	2.9	-1.427	-0.394
1000	0.2	0.5	<i>n.a.</i>	<i>n.a.</i>	3226.7	5.5	-1.382	-0.383
1000	0.4	0.5	<i>n.a.</i>	<i>n.a.</i>	2433.5	3.8	-1.450	-0.406
1000	0.8	0.5	<i>n.a.</i>	<i>n.a.</i>	1808.0	2.3	-1.456	-0.409
100	0.2	0.2	<i>n.a.</i>	<i>n.a.</i>	5390.5	9.4	-1.359	-0.384
100	0.2	0.4	<i>n.a.</i>	<i>n.a.</i>	3820.5	7.2	-1.360	-0.367
100	0.2	0.6	<i>n.a.</i>	<i>n.a.</i>	3148.2	6.4	-1.344	-0.360
100	0.2	0.8	<i>n.a.</i>	<i>n.a.</i>	2758.9	5.4	-1.339	-0.363
100	0.2	1.0	<i>n.a.</i>	<i>n.a.</i>	2499.6	5.2	-1.317	-0.353
100	0.2	0.5	0.05	2	4623.6	6.7	-1.423	-0.399
100	0.2	0.5	0.10	10	5430.8	6.4	-1.462	-0.428
100	0.2	0.5	0.20	10	7043.1	6.8	-1.454	-0.442
100	0.2	0.5	0.40	10	10910.1	11.9	-1.387	-0.448

Theoretische Überlegungen für den erwarteten Exponenten des Allelfrequenz-Spektrums nach einer Range-Expansion präsentiere ich in ABSCHN. B.2. Dabei wird der Exponent des Potenzgesetzes von der Form der Expansionsfront beeinflusst. So erwartet man für gerade Fronten einen Exponenten von $a = -1.33$ und für wellige Fronten einen Exponenten von $a = -1.4$. Die in den Simulationen gemessenen Exponenten liegen in diesem Bereich. Eine Änderung der Parameterwerte hat meinen Daten zufolge keinen essentiellen Einfluss auf den Exponenten (siehe TAB. 3.3). Dennoch ist zu erkennen, dass mit steigender Migrationsrate m , steigender Kapazitätsgrenze K sowie einem stärkeren Allee-Effekt der Exponent kleiner wird. Mit steigender Wachstumsrate r wird der Exponent hingegen größer und liegt nahe $a = -1.33$ (siehe TAB. 3.3).

Abgesehen vom Potenzgesetz findet sich in den Spektren auch ein Bereich hoher Allelfrequenzen, in dem die Häufigkeitsdichte mit steigender Frequenz p_f nicht geringer wird (Plateau in ABB. B.9, ABB. 3.7 und ABB. B.8 rechts der gestrichelten Linie). In diesem Bereich finden sich solche Simulationen, in denen die Frequenz des abgeleiteten Allels noch weiter steigen könnte, würde die Range-Expansion länger dauern. Die Anzahl der Simulationen in diesem Bereich ist ein Maß für die Surfing-Wahrscheinlichkeit. Für einige Parameterwerte werden in wenigen Simulationen sehr hohe Frequenzen erreicht. Für andere Parameterwerte werden hingegen mehr Simulationen durch Gene-Surfing beeinflusst, die dabei erreichten Frequenzen sind allerdings geringer.

Wie Klopstein et al. (2006) bereits mit einer vergleichbaren Analyse zeigen konnten, ist auch in meinen Simulationen die Surfing-Wahrscheinlichkeit positiv mit der Wachstumsrate r und negativ mit der Migrationsrate m korreliert (siehe ABB. B.8 und ABB. 3.7). Allerdings deuten meine Ergebnisse darauf hin, dass die Surfing-Wahrscheinlichkeit mit steigender Kapazitätsgrenze K zunimmt (ABB. B.9). Klopstein et al. (2006, Tab. 1 und Fig. 3) haben hier hingegen eine negative Korrelation gefunden. In Simulationen, in denen die Population unter dem Einfluss eines Allee-Effektes wächst, steigt die Wahrscheinlichkeit, durch Gene-Surfing hohe Frequenzen zu erreichen, ebenfalls (ABB. 3.8).

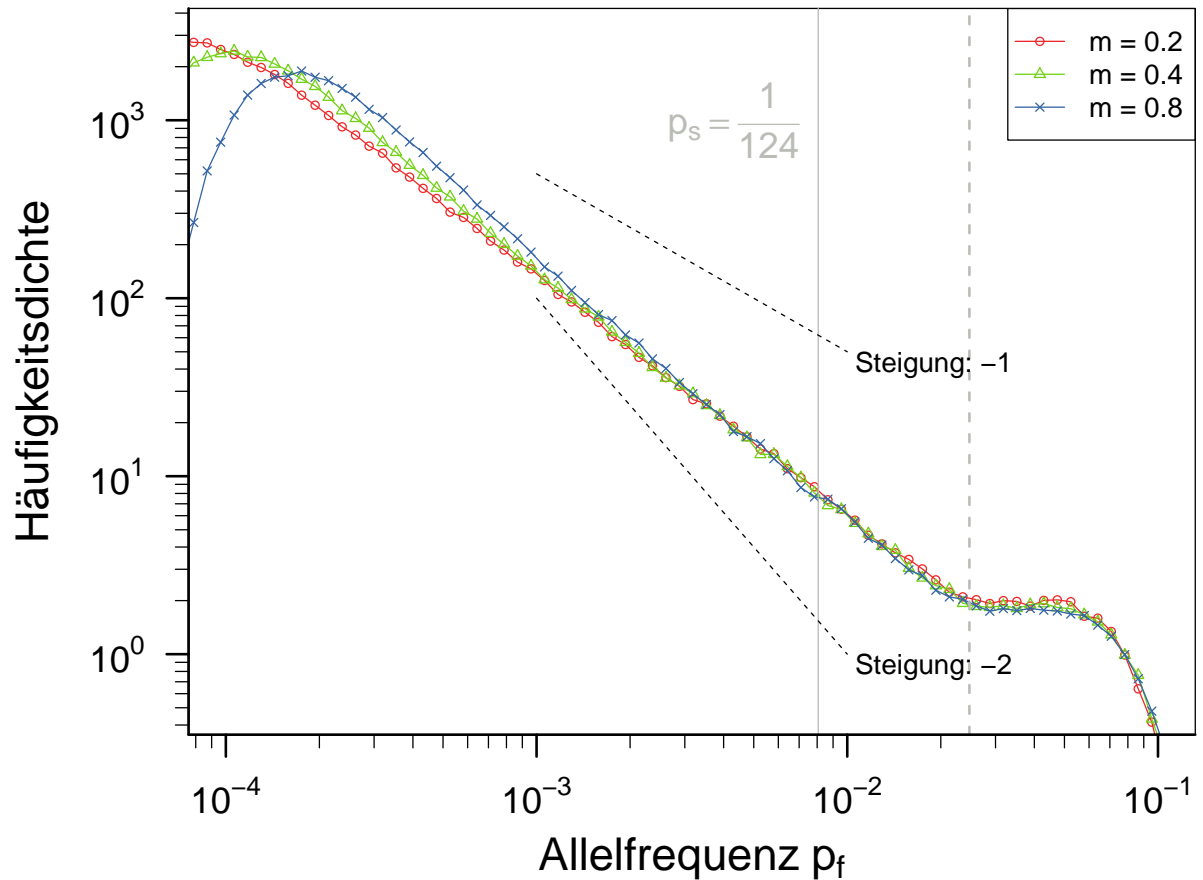


Abbildung 3.7.: **Allelfrequenz-Spektren aus Simulationen von Range-Expansions mit verschiedenen Migrationsraten m .** In der doppelt-logarithmischen Darstellung ist eine Gerade zu erkennen, was auf ein Potenzgesetz hindeutet. Der Exponent dieses Potenzgesetzes entspricht der Steigung der Geraden und lässt sich durch lineare Regression ermitteln (siehe [ABB. B.11](#) und [ABB. B.12](#)). Mit zunehmender Migrationsrate sinken sowohl der gemessene Exponent ([TAB. 3.3](#)) als auch die Zahl der Simulationen, in denen bei einer längeren Dauer noch höhere Frequenzen durch Gene-Surfing erreicht werden könnten (Anzahl von Simulationen rechts der gestrichelten vertikalen Linie). Die Daten stammen aus Simulationen für Standing-genetic-Variation ($p_s = 0.00806$; durchgezogene vertikale Linie) auf der Großen Welt ($w = 1000$; $h = 1000$; [ABB. A.4](#)). Gezeigt sind Ergebnisse aus Simulationen mit einer Kapazitätsgrenze von $K = 100$ und einer Wachstumsrate von $r = 0.5$. Die beiden diagonalen Linien zeigen Referenzgeraden mit den Steigungen $a = -1$ und $a = -2$.

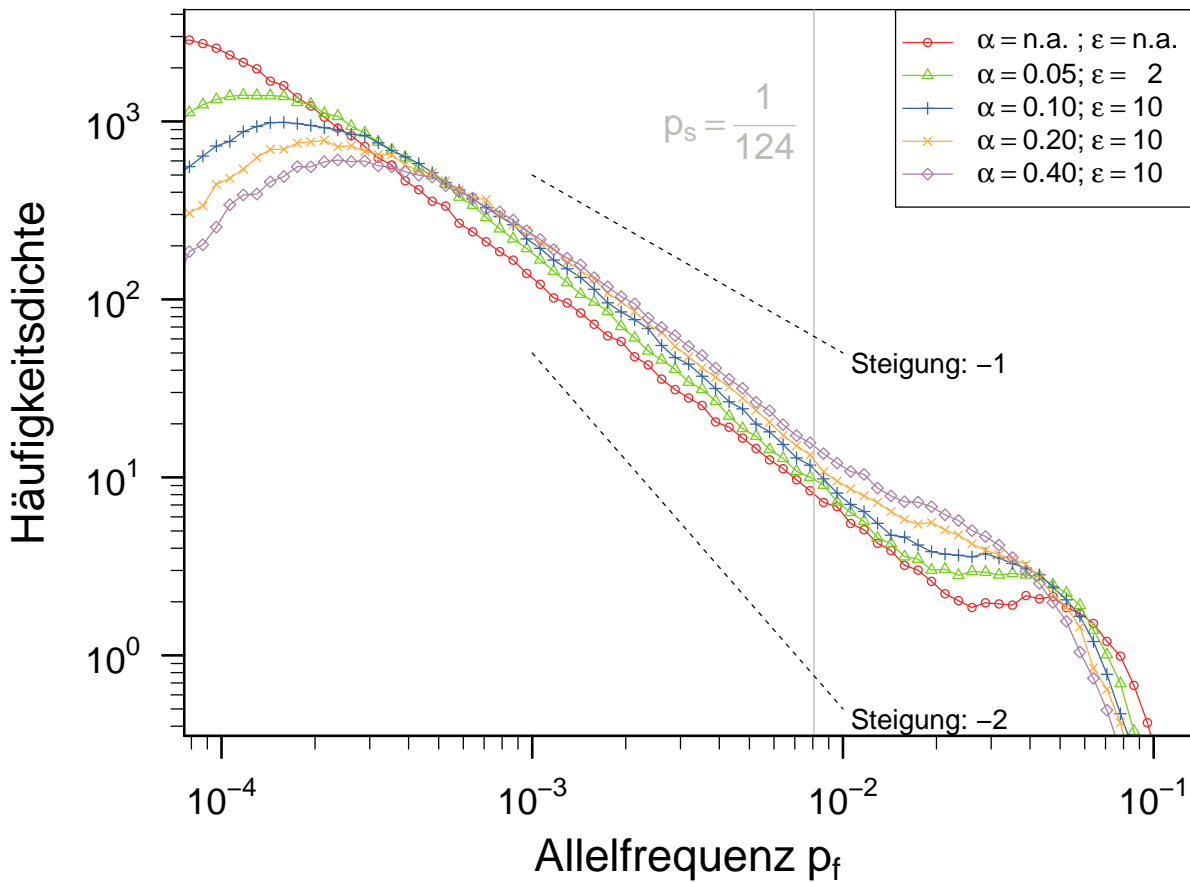


Abbildung 3.8.: **Allelfrequenz-Spektren aus Simulationen von Range-Expansions mit verschiedenen starken Allee-Effekten.** Beim Allee-Effekt ist die Wachstumsrate unterhalb einer Populationsdichte α stark verringert (siehe TEILABSCHN. A.2.3), wodurch sich die Front deutlich langsamer bewegt als ohne einen Allee-Effekt (TAB. 3.3). In Simulationen mit einem Allee-Effekt kommt es häufiger vor, dass das abgeleitete Allel eine Frequenz p_f erreicht, die größer als die Startfrequenz p_s ist. Der Allee-Effekt führt demnach dazu, dass Mutationen mehr Chancen bekommen, den Kontakt zur Front zu behalten und zu surfen als in Simulationen ohne einen Allee-Effekt. Die maximal erreichte Frequenz in den Spektren aus Simulationen mit einem Allee-Effekt ist allerdings geringer. Für verschiedene Allee-Effekte ist in der doppelt-logarithmischen Darstellung ein Potenzgesetz erkennbar, dessen Exponent im Bereich von $a_k = -1.4$ liegt (siehe TAB. 3.3 und vergleiche ABB. 3.7). Die Daten stammen aus Simulationen für Standing-genetic-Variation ($p_s = 0.00806$; durchgezogene vertikale Linie) auf der Großen Welt ($w = 1000$; $h = 1000$; ABB. A.4). Gezeigt sind Ergebnisse aus Simulationen mit einer Kapazitätsgrenze von $K = 100$, einer Wachstumsrate von $r = 0.5$ und einer Migrationsrate von $m = 0.2$. Die beiden diagonalen Linien zeigen Referenzgeraden mit den Steigungen $a = -1$ und $a = -2$.

4. Diskussion

Range-Expansions sind ein häufiges Phänomen in der belebten Umwelt. Sie können z.B. in Folge von Klimaveränderungen, der Verschleppung von Arten oder im Anschluss an die Entstehung einer neuen Art auftreten. Aufgrund von Range-Expansions kann sich die genetische Zusammensetzung innerhalb von Populationen stark verändern (Edmonds et al., 2004; Nielsen et al., 2007). In dieser Arbeit habe ich den Einfluss von Range-Expansions auf Allelfrequenz-Spektren untersucht. Viele genetische Varianten haben vor Beginn einer Range-Expansion typischerweise geringe Frequenzen (siehe ABB. B.14). Daher wurde von mir insbesondere das Schicksal solcher Allele betrachtet.

Dabei habe ich herausgefunden, dass die resultierenden Allelfrequenz-Spektren durch ein Potenzgesetz beschrieben werden können. Der Exponent ist spezifisch für Range-Expansions und von der Beschaffenheit der Front abhängig. Den Exponenten aus meinen räumlichen Simulationen vergleiche ich in ABSCHN. 4.1 mit den Erwartungen für nicht räumliche Populationen. Die Parameter der Range-Expansion haben keinen wesentlichen Einfluss auf diesen Exponenten, wirken sich allerdings auf die Surfing-Wahrscheinlichkeit aus.

Die Auswirkungen von Range-Expansions auf das Schicksal neu auftretender Mutationen sind vergleichbar mit den Auswirkungen auf bestehende genetische Variationen. Daher diskutiere ich in ABSCHN. 4.2 die relative Wichtigkeit beider Typen von genetischer Variation für die Allelfrequenz-Spektren. Dabei gehe ich auch auf die Auswirkung räumlicher Hindernisse auf die Allelfrequenz-Spektren ein.

In ABSCHN. 4.3 stelle ich dar, welche Bedeutung die Ergebnisse für die Anwendbarkeit auf echte genetische Daten haben. Darüber hinaus diskutiere ich, wie gut das von mir verwendete Modell die Realität abbilden kann und gehe im Anschluss auf Erweiterungsmöglichkeiten des Modells ein (ABSCHN. 4.4).

4.1. Range-Expansions generieren Allelfrequenz-Spektren, die einem Potenzgesetz folgen

In meinem Modell ist die Wahrscheinlichkeit dafür, dass eine Mutation an der Front zu finden ist, durch die Mutationsrate μ bzw. die Anfangsfrequenz p_s gegeben. Ausgehend von dieser Wahrscheinlichkeit habe ich die Verteilung der Allelfrequenzen nach Simulationen von verschiedenen demographischen Szenarien betrachtet (10^6 Realisationen pro Parametersatz). In diesen Allelfrequenz-Spektren $f(p_f)$ finde ich nach einer Range-Expansion eine sehr große Bandbreite an gemessenen Frequenzen p_f , die in den anderen demographischen Szenarien nach der selben Zeit nicht zu erkennen ist (siehe z.B. ABB. 3.2). Dieses breite Spektrum an Allelfrequenzen hatten bereits [Edmonds et al. \(2004\)](#) erwähnt. Mit meinen Simulationen konnte ich zeigen, dass die Verteilung der Frequenzen p_f einem Potenzgesetz der Form $f(p_f) = C \cdot p_f^\alpha$ folgt. Potenzgesetze haben typischerweise eine große Varianz in den Daten ([Newman, 2005](#)), was die große Standardabweichung bei den Allelfrequenzen in der Studie von [Edmonds et al. \(2004, Table 2\)](#) erklärt.

Der Exponent des Potenzgesetzes ist von der Form der Front abhängig

Potenzgesetze finden sich in einer Vielzahl empirischer Daten ([Clauset et al., 2009](#); [Newman, 2005](#)). Diverse Mechanismen, die zu diesen Verteilungen führen können, sind in der Literatur beschrieben ([Newman, 2005](#)). So sind z.B. viele Charakteristika von Zufallsbewegungen (*random walks*) gemäß einem Potenzgesetz verteilt ([Newman, 2005](#)). Mutationen, die an der Front bestand haben, können Sektoren bilden, deren Ränder sich ebenfalls wie Random-Walker verhalten (siehe ABB. B.13). Die Überlebenszeiten dieser Sektoren an der Front sind dabei gemäß einem Potenzgesetz verteilt ([Newman, 2005](#)). In ABSCHN. B.2 nutze ich diese Abhängigkeit, um die erwartete Größenverteilung der Sektorflächen als Maß für die Frequenz abzuschätzen. Dabei ist zu erkennen, dass der Exponent der Verteilung von der Struktur der Front abhängt. Grund hierfür ist, dass sich die Sektorgrenzen bei ungleichmäßigen Fronten superdiffusiv bewegen ([Hallatschek et al., 2007](#)) und es dadurch schneller zu einem Zusammentreffen der Random-Walker kommen kann.

Der Exponent des Potenzgesetzes ist spezifisch für Range-Expansions

Range-Expansions führen sowohl bei Standing-genetic-Variation als auch bei Ongoing-Mutations im Bereich hoher Frequenzen zu qualitativ identischen Allelfrequenz-Spektren p_f : Sie haben den gleichen Exponenten (Plateau in ABB. 3.4). In diesem Bereich befinden sich diejenigen Allele, die in Simulation von SV und OM gleichermaßen gesurft sind. Dabei ist die Surfing-Wahrscheinlichkeit proportional zur Anfangsfrequenz bzw. der Mutationsrate (ABB. 3.3). Diese Verschiebung entlang der Abszisse findet sich im Faktor C des Potenzgesetzes wieder (GL. 3.5).

Meine Ergebnisse aus Simulationen auf zweidimensionalen Gittern vergleiche ich nun mit stabilen und wachsenden, durchmischten (*well-mixed*) Populationen. Für Allelfrequenz-Spektren einer durchmischten Population konstanter Größe erwartet man in der doppelt-logarithmischen Darstellung eine Steigung von $a = -1$ (Wakeley, 2008). Im Gegensatz dazu sind Allelfrequenzen einer durchmischten, exponentiell wachsenden Population so verteilt, dass sie einen Exponenten von $a = -2$ zeigen (Durrett, 2013). Den gleichen Exponenten erwartet man auch in der Verteilung der Allelfrequenzen schnell adaptierender Populationen (Neher and Hallatschek, 2013).

In Verteilungen mit einem kleinen Exponenten sind niedrige Frequenzen stärker repräsentiert. Daher erwartet man in exponentiell wachsenden Populationen einen erhöhten Anteil niedriger Frequenzen. Für Menschen, deren Zahl in den letzten zehntausend Jahren vermutlich exponentiell gewachsen ist, wurde eine solche Häufung an seltenen genetischen Varianten bereits gefunden (Keinan and Clark, 2012). Der von mir gemessene Exponent für räumliches Wachstum liegt genau zwischen den erwarteten Exponenten einer stabilen und einer exponentiell wachsenden Population. Dementsprechend können auch Range-Expansions einen Anstieg seltener Allele erklären. Dem Stand der Forschung zufolge haben sowohl Range-Expansions als auch exponentielles Wachstum in der Geschichte des Menschen stattgefunden. Daher bleibt zu klären, welche Verteilungen nach einer Kombination beider Szenarien zu erwarten wären.

Das Potenzgesetz ist robust gegenüber dem Allee-Effekt und weiteren Parametern der Simulation

In den Simulationen habe ich Exponenten zwischen $a = -1.46$ und $a = -1.32$ gemessen. Diese Werte liegen im theoretisch erwarteten Bereich für geradlinige ($a = -1.33$) und

4. Diskussion

wellige ($a = -1.4$) Frontverläufe (siehe ABSCHN. B.2). In meinen Simulationen konnten sich die Individuen komplett frei bewegen (*phenotype limited range expansion*), was in der Regel zu ungleichmäßigen Frontverläufen und somit zu Exponenten $a > 1.33$ führt. Für Simulationen, deren Geschwindigkeit z.B. durch einen langsam schmelzenden Gletscher begrenzt ist (*boundary limited range expansion*; [Nullmeier and Hallatschek \(2013\)](#)) erwarte ich hingegen Exponenten, die bei $a \approx 1.33$ liegen.

Wie bereits [Klopfstein et al. \(2006\)](#) finde auch ich Gene-Surfing für ein breites Spektrum von Parameterwerten. Durch eine Änderung der Parameterwerte (z.B. Stärke des Wachstums, Kapazitätsgrenzen, ...) ändern sich auch die Exponenten des Potenzgesetzes leicht. Allerdings liegen diese Schwankungen im Bereich der Abweichungen, die ich zwischen den beiden Messmethoden finde. Daher kann ein grundlegender Einfluss der Parameterwerte auf die Steigung nicht zweifelsfrei festgestellt werden. Da sich insbesondere bei den Spektren des Allee-Effektes die Messung des Exponenten als schwierig herausstellte, sollten für weitere Analysen alternative Messmethoden angewendet werden ([Clauset et al., 2009](#)).

Der von mir simulierte Allee-Effekt hat erwartungsgemäß den Grad der Inzucht und damit den Verlust genetischer Vielfalt an der Front verringert (siehe folgender Abschnitt). Die Sektorbildung, mit deren Hilfe ich in ABSCHN. B.2 die Form der Allelfrequenz-Spektren erkläre, wird dadurch allerdings nicht verhindert. Einen grundlegenden Einfluss auf den Exponenten der Verteilung konnte ich daher nicht feststellen.

Die Parameter einer Range-Expansion beeinflussen die Surfing-Wahrscheinlichkeit

Den Einfluss der Parameterwerte auf den Exponenten des Potenzgesetzes konnte ich nicht zweifelsfrei bestimmen. Allerdings finde ich, dass eine steigende Wachstumsrate r die Anzahl der Surfing-Ereignisse erhöht (siehe ABB. B.8). Den Grund sehe ich darin, dass zwei Formen der genetischen Drift während einer Range-Expansion eine Rolle spielen. Zum einen kann die genetische Vielfalt durch die zufällige Auswahl einer geringen Anzahl von Individuen während der Besiedlung verloren gehen (*Gründereffekt*). Zum anderen gibt es innerhalb der Deme genetische Drift beim Übergang von einer Generation zur nächsten (siehe ABB. A.1). Die Zeit, in der im Mittel ein (seltenes) Allel aufgrund genetischer Drift verloren geht, ist von der Größe der lokalen Population abhängig. Sie ist dabei um so kürzer, je kleiner die Population ist ([Hartl et al., 2007](#); [Kimura and Ohta,](#)

4.1. Range-Expansions generieren Allelfrequenz-Spektren, die einem Potenzgesetz folgen

1969). Ist also ein seltenes Allel an einer Besiedlung beteiligt und bleibt die Population über einen längeren Zeitraum klein, so geht diese Mutation schnell verloren und hat keine Chance mehr, vom Surfing-Effekt betroffen zu sein. Wächst die Population hingegen schnell, so hat jedes Individuum zum einen die Chance auf mehr Nachkommen, zum anderen verlängert sich aufgrund der gestiegenen Populationsgröße die Wartezeit bis zum Verlust der Mutation. In dieser Zeit besteht die Möglichkeit, dass die Mutation für weitere Besiedlungen ausgewählt wird.

Eine positive Korrelation für die Surfing-Wahrscheinlichkeit finde ich auch bei der Kapazitätsgrenze K (siehe [ABB. B.9](#)). Den Grund sehe ich darin, dass bei der Neubesiedlung jedes Doms immer bis zu einem gewissen Anteil seiner Kapazitätsgrenze besiedelt wird. Die absolute Individuenzahl bei der Besiedlung steigt daher mit zunehmender Kapazitätsgrenze, wodurch die genetische Variation im Mittel über einen längeren Zeitraum an der Front erhalten bleibt ([Hartl et al., 2007](#); [Kimura and Ohta, 1969](#)). Dadurch bekommen auch seltene Allele mehr Möglichkeiten, für weitere Besiedlungen ausgewählt zu werden und in Folge dessen zu hohen Frequenzen zu surfen. Im Gegensatz zu meinen Ergebnissen hatten [Klopfstein et al. \(2006\)](#) eine negative Korrelation zwischen Kapazitätsgrenze und Surfing-Wahrscheinlichkeit gefunden. Den Grund sehe ich darin, dass [Klopfstein et al.](#) in ihren Simulationen ein einzelnes Individuum bei der Erstbesiedlung mutiert haben. Wie bereits erwähnt, steigt die absolute Zahl der Individuen bei der Besiedlung mit zunehmender Kapazitätsgrenze. Dadurch sinkt allerdings der Anteil (Frequenz), den das mutierte Individuum an der Gründerpopulation ausmacht. Wie ich zuvor in [ABB. 3.3](#) gezeigt habe ist die Surfing-Wahrscheinlichkeit aber proportional zur Frequenz und muss daher zwangsläufig bei diesem Simulationsdesign sinken.

Wie bereits durch [Taylor and Hastings \(2005\)](#) beschrieben verlangsamt ein Allee-Effekt die Ausbreitung und das Wachstum an der Front. Infolgedessen steigt die absolute Zahl der Migrationsereignisse und erlaubt es Mutationen, die den Kontakt zur Front verloren hatten, erneut an die Front zu gelangen. Es kommt also in den Simulationen mit einem Allee-Effekt in der Wachstumsfunktion zum Erhalt eines größeren Teils der genetischen Vielfalt an der Front und damit zu einer erhöhten Surfing-Wahrscheinlichkeit (siehe [ABB. 3.8](#)).

Mit zunehmender Migrationsrate m nimmt die Breite des noch nicht voll besiedelten Bereiches zu, wodurch die Surfing-Wahrscheinlichkeit sinkt. Zwar findet auch hier wie beim Allee-Effekt ein stärkerer Austausch von Individuen statt, im Gegensatz zum Allee-Effekt wächst die Population allerdings bereits bei geringen Populationsdichten. Schaffen

4. Diskussion

es wenige Individuen eines Typs, ein Dem zu besiedeln, so können sie sofort wachsen. Dadurch verringern sie das mögliche Wachstum späterer Immigranten und verschaffen ihren Nachkommen auf diese Weise einen Vorteil für die Besiedlung neuer Gebiete (*high density blocking*).

In Allelfrequenz-Spektren mit hoher Surfing-Wahrscheinlichkeit findet man typischerweise geringere maximale Frequenzen als in Spektren mit geringerer Surfing-Wahrscheinlichkeit. Das stützt meine Erwartung, da die Fixierungswahrscheinlichkeit eines neutralen Allels der Anfangsfrequenz p_s entspricht (Hartl et al., 2007). Daher muss auch der Mittelwert aller Allelfrequenzen im Spektrum der Anfangsfrequenz p_s entsprechen.

4.2. Mutationen aus der Ursprungspopulation dominieren Allelfrequenz-Spektren

In dieser Arbeit habe ich mir vor allem den Bereich hoher Allelfrequenzen in den Spektren angeschaut, da hier die „gesurften“ Allele zu finden sind. Damit Allele surfen können, müssen sie sich nahe der Front befinden (Klopfstein et al., 2006). Betrachten wir zunächst die Wahrscheinlichkeit dafür, dass sich eine Mutation in einem zufällig ausgewählten Individuum an der Front befindet. Diese Wahrscheinlichkeit entspricht bei Standing-genetic-Variation in der ersten Generation genau der Anfangsfrequenz p_s dieser Mutation. Bei Ongoing-Mutations können Mutationen allerdings in jeder Generation mit einer Mutationsrate μ in jedem Individuum entstehen. Ob vor allem neu auftretende Mutationen oder Mutationen aus der Ursprungspopulation vom Gene-Surfing betroffen und somit im Bereich hoher Allelfrequenzen zu finden sind, war bisher ungeklärt (Travis et al., 2007).

Zunächst konnte ich zeigen, dass die Surfing-Wahrscheinlichkeit proportional zu μ bzw. p_s ist (ABB. 3.3). Im Anschluss habe ich den relativen Beitrag (ausgedrückt durch Γ) von Ongoing-Mutations und Standing-genetic-Variation zum Allelfrequenz-Spektrum bewertet. Bei realistischen Werten für μ und die Dauer t_e der Range-Expansion habe ich festgestellt, dass ein Großteil der gesurften Allele bereits in der Ursprungspopulation vorhanden war (siehe ABSCHN. 3.3). Allerdings steigt die Bedeutung von Ongoing-Mutations für die Allelfrequenz-Spektren (linear) mit zunehmender Dauer t_e einer Range-Expansion (ABB. B.5). Der Grund dafür ist, dass es bei längeren Range-Expansions mehr Gelegenheiten gibt, in denen Mutationen an der Front auftreten und durch Gene-Surfing beeinflusst werden können. Die relative Bedeutung Γ von Ongoing-Mutations ist

4.2. Mutationen aus der Ursprungspopulation dominieren Allelfrequenz-Spektren

proportional zur Laufzeit t_e und liegt in deren Größenordnung (Steigung in ABB. B.5 ist ≈ 1 ; $t_e \approx \Gamma$). Nun kann man diejenige Dauer t_e einer Range-Expansion bestimmen, nach der *neue Mutationen* häufiger im Bereich hoher Frequenzen zu finden sind als Mutationen, die in der Ursprungspopulation eine Frequenz von p_s hatten. Diese Dauer ergibt sich direkt aus μ und p_s (vergleiche auch TAB. 3.1):

$$t_e \approx \Gamma > \frac{p_s}{\mu}.$$

In afrikanischen Genomen gibt es mehr als 10^7 Genorte, an denen die Frequenz p_f des abgeleiteten Allels in der Größenordnung von $p_f = 10^{-2}$ liegt (siehe ABB. B.14). Nimmt man diese Frequenz als mittlere Anfangsfrequenz p_s , so müsste eine Range-Expansion bei einer Mutationsrate von $\mu = 10^{-8}$ (1000 Genomes Project Consortium et al., 2012; Conrad et al., 2011) in etwa $t_e = 10^6$ Generationen dauern, bis Ongoing-Mutations den gleichen Einfluss auf Allelfrequenz-Spektren haben wie Standing-genetic-Variation. Für den Menschen sind allerdings Range-Expansions mit einer Dauer von $t_e = 10^3$ Generationen realistischer (Oppenheimer, 2012; Currat and Excoffier, 2005).

Geographische Hindernisse beeinflussen vor allem Mutationen aus Standing-genetic-Variation

Wie bereits im vorherigen Abschnitt erwähnt ist ein entscheidender Unterschied zwischen Standing-genetic-Variation (SV) und Ongoing-Mutations (OM) der Zeitpunkt, zu dem die Mutationen (relativ zur Range-Expansion) in der Population vorhanden sind. Bei SV sind die Mutationen alle schon zu Beginn der Range-Expansion in der Population vorhanden. Schafft es eine Mutation aus der Ursprungspopulation durch eine geographische Engstelle, so ist es sehr wahrscheinlich, dass sie eine hohe Frequenz erreicht. Schafft sie es hingegen nicht durch die Engstelle, so bleibt ihre Frequenz (insbesondere bei frühen Hindernissen) gering (ABB. 3.5). Das deckt sich mit den Erwartungen, die aus den Ergebnissen von Burton and Travis (2008) folgen.

Im Gegensatz dazu können OM während der Range-Expansion in jeder Generation mit der Wahrscheinlichkeit μ (Mutationsrate) auftreten. Hindernisse haben dabei keinen Einfluss darauf, ob eine Mutation entsteht oder nicht. Ist also eine genetische Variante aufgrund einer geographischen Engstelle verloren gegangen, besteht weiterhin die Möglichkeit, dass die Mutation im Anschluss erneut an der Front auftritt. Daher ist in den

4. Diskussion

Simulationen von Ongoing-Mutations keine Aufteilung in „entweder hohe oder niedrige Frequenzen“ zu erkennen (siehe [ABB. 3.5](#)).

In meinem Modell habe ich nachgestellt, dass die Populationen gleichmäßig über den Raum verteilt wachsen. Allerdings ist auch eine Ausbreitung entlang bestimmter Wanderrouen oder in Trecks denkbar. [ABB. B.10](#) zeigt Allelfrequenz-Spektren aus Simulationen auf einer langgezogenen Welt. Vergleichbar zu den Ergebnissen für geographische Engstellen kommt es auch hier zu einer sichtbaren Aufteilung in Simulationen, die sich nahe der Fixierung befinden und solcher Simulationen, in denen das abgeleitete Allel selten ist. Daher kann das Verhältnis aus Breite und Lauflänge ebenfalls als Engstelle aufgefasst werden.

In Simulationen auf der Europakarte (siehe [ABB. 3.6](#)) ist zu erkennen, dass die relative Bedeutung von Ongoing-Mutations für die Allelfrequenz-Spektren steigt. Die Ursache sehe ich darin, dass nach der Passage einer Engstelle die Population entlang einer gebogenen und somit längeren Front wächst (siehe z.B. [ABB. 1.4](#)). Dadurch ergeben sich mehr Möglichkeiten für Mutationen, an der Front aufzutreten und im Anschluss zu surfen. Da aller Voraussicht nach auch räumliche Inhomogenitäten zu gebogenen Fronten führen ([Wegmann et al., 2006](#)), sollte deren quantitativer Einfluss eingehender untersucht werden.

Geographische Hindernisse haben mit Sicherheit die Ausbreitung der Menschen beeinflusst. Sie wirken sich vor allem auf genetische Variationen aus, die bereits in der Ursprungspopulation vorhanden waren. Um eventuelle Spuren kürzlicher Range-Expansions in den Allelfrequenz-Spektren echter genetischer Daten zu finden, sollten daher die Spektren für Ongoing-Mutations und Standing-genetic-Variation wenn möglich getrennt voneinander betrachtet werden.

4.3. Anwendung auf echte genetische Daten

In dieser Arbeit habe ich die Auswirkungen von Range-Expansions auf Allelfrequenz-Spektren untersucht. Allelfrequenz-Spektren lassen sich auch für echte genetische Daten erzeugen, indem man für viele Genorte die Frequenz des abgeleiteten Allels misst. Das 1000-Genome-Projekt stellt solche Daten für menschliche Populationen zur Verfügung ([1000 Genomes Project Consortium et al., 2012](#); [1000 Genomes Project Consortium, 2010](#)). Nehmen wir zunächst an, dass alle $n \approx 3.3 \times 10^9$ Genorte des menschlichen Genoms unabhängig voneinander vererbt werden. In diesem Falle entsprechen die Allelfrequenz-Spektren echter genetischer Daten konzeptionell denen aus meinen Simulationsdaten, die

ich aus jeweils 10^6 unabhängigen Simulationen eines einzelnen Genortes erstellt habe. Tatsächlich sind Genorte mit ihren genetischen Varianten dadurch aneinander gebunden, dass sie gemeinsam vererbt werden (*linkage*). Wäre nur ein einziges haploides Individuum und dessen Nachkommen für eine Neubesiedlung verantwortlich, so würden alle Varianten dieses Individuums fixiert werden. In dem Falle wäre das Allelfrequenz-Spektrum der abgeleiteten Allele bimodal mit Peaks bei $p_f = 1$ und $p_f = 0$.

Allerdings werden Genorte nicht immer gemeinsam vererbt, sondern können durch *Crossing-over* während der Meiose untereinander neu kombiniert werden (Ardlie et al., 2002). Sind zwei Genorte einen Centimorgan (cM) voneinander entfernt, so werden sie im Schnitt nach 100 Generationen durch Rekombination getrennt. Im menschlichen Genom entspricht $1cM \approx 7.5 \times 10^5 bp$ (bp = Basenpaare) (Lodish et al., 2003). Daher erwartet man im Mittel $3.3 \times 10^9 / 7.5 \times 10^7 = 44$ Rekombinationsereignisse pro Generation (siehe auch Hassold et al., 2000, und darin enthaltene Referenzen). Genorte, für die mehr als ein Allel im menschlichen Genom bekannt ist, liegen im Mittel weniger als $1000bp$ auseinander (siehe ABB. B.14) und werden daher selten durch Rekombination getrennt. Mutationen, die aus Standing-genetic-Variation stammen, sollten daher häufiger gemeinsam von Gene-Surfing betroffen sein. Welche Auswirkungen dieses gemeinsame Surfing auf die Allelfrequenz-Spektren hat, ist bisher unbekannt und kann Gegenstand zukünftiger Forschung sein.

Bei Ongoing-Mutations erwartet man, dass pro Individuum und Generation etwa $n \cdot \mu = 3.3 \times 10^9 \cdot 10^{-8} = 33$ Mutationen auftreten (Conrad et al., 2011). Im Gegensatz zu Mutationen aus Standing-genetic-Variation treten diese neuen Mutationen im Mittel in Bereichen auf, die unabhängig voneinander vererbt werden. Daher erwarte ich, dass mögliche Auswirkungen einer gemeinsamen Vererbung einen geringeren Einfluss auf die Allelfrequenz-Spektren von Standing-genetic-Variation haben.

In ABB. B.14 und ABB. B.15 zeige ich Spektren aus Frequenzen abgeleiteter Allele für vier Populationen des Menschen (siehe auch ABB. A.7). Für Allele, die vermutlich aus Standing-genetic-Variation stammen (ABB. B.14), finde ich in den Allelfrequenz-Spektren dreier Populationen (Europa, Amerika, Asien) eine ähnliche Steigung. Das kann als Indiz dafür gewertet werden, dass gleiche Prozesse die Spektren geformt haben, insbesondere weil diese Populationen nach aktuellem Stand der Wissenschaft ihr Verbreitungsgebiet (ausgehend von Afrika) vergrößert haben. Was tatsächlich passiert ist, konnte hier allerdings nicht festgestellt werden und sollte Gegenstand zukünftiger Forschung sein.

Allelfrequenz-Spektren von Standing-genetic-Variation können durch geographische

4. Diskussion

Hindernisse und Kopplung der Genorte (*linkage*) beeinflusst werden. Daher habe ich in [ABB. B.15](#) aus Daten des 1000-Genome-Projektes Spektren solcher abgeleiteter Allele erstellt, die vermutlich während der Range-Expansion aufgetreten sind. Dabei ist zu erkennen, dass es einen Bereich von Frequenzen gibt ($p_f \approx 0.5$), die an keinem Genort erreicht werden. Hier gibt es Ähnlichkeiten zu Allelfrequenz-Spektren von Standing-genetic-Variation, die durch geographische Hindernisse beeinflusst wurden (siehe [ABB. B.10](#) und [ABB. 3.5](#)). Zum einen ist aber eine Auswirkung räumlicher Hindernisse bei Ongoing-Mutations nicht zu erwarten, zum anderen sollte ein Einfluss der Geographie auch in den Spektren von Standing-genetic-Variation zu erkennen sein.

Es ist daher wahrscheinlich, dass die hochfrequenten Allele nicht aus Ongoing-Mutations stammen, sondern aufgrund der geringen Stichprobengröße ($n = 246$; siehe [ABB. A.7](#)) nicht in afrikanischen Individuen sequenziert werden konnten. Alternativ können die Allele auch in Afrika verloren gegangen sein. Eine weitere Erklärung können sogenannte *Range-Constrictions* (Verkleinerungen des Verbreitungsraumes) und *Range-Shifts* (Verschiebungen des Verbreitungsraumes) geben ([Arenas et al., 2012](#)). So können Allele tatsächlich spezifisch für eine Population sein. Der Grund für ihre Häufigkeit ist allerdings nicht eine einzige Range-Expansion, sondern eine Folge von Range-Expansions und Range-Constrictions.

Um den für Range-Expansions spezifischen Exponenten in echten Daten messen zu können, müsste man den Bereich niedriger Allelfrequenzen betrachten. Allerdings wurden im 1000-Genome-Projekt weniger als 1000 Individuen pro Population sequenziert ([1000 Genomes Project Consortium et al., 2012](#)). Daher können nur Frequenzen von $p_f > 10^{-3}$ gemessen werden, wodurch eine exakte Messung nicht möglich ist. Zukünftige Sequenzierprojekte mit Stichprobengrößen von $n > 10^5$ werden solche Analysen aller Voraussicht nach ermöglichen.

4.4. Ausblick

Simulationen können nur einen Teil der Prozesse und Szenarien in der Natur abbilden. Einige von denen (Migration, Wachstum, geographische Hindernisse, Allee-Effekt, ...) habe ich in meinen Modellen berücksichtigt. Es gibt allerdings eine Reihe weiterer Prozesse, die einen Einfluss auf die genetische Vielfalt während einer Range-Expansion haben können. So konnten [Fayard et al. \(2009\)](#) zeigen, dass Langstreckenmigration (*Long distance dispersal*; LDD) die Häufigkeit von Surfing-Ereignissen herabsetzt. Diese

Form der Ausbreitung gibt es bei Pflanzen und Tieren ([Waters et al., 2013](#)) und zeigt sich auch im Reiseverhalten des heutigen Menschen ([Brockmann et al., 2006](#)). Eine Population mit einer hohen Rate an Langstreckenmigration entspricht in etwa einer wachsenden, durchmischten Population. Daher erwarte ich für den Exponenten a des Allelfrequenz-Spektrums einer Population mit Langstreckenmigration einen Wert im Bereich von $a = -2$.

In dieser Arbeit habe ich den Einfluss von Selektion nicht betrachtet, da die meisten Mutationen im menschlichen Genom neutral oder nahezu neutral sind ([Kimura, 1968](#)). Dennoch sollten die Auswirkungen von Selektion zukünftig betrachtet werden, da auch sie Spuren in Allelfrequenz-Spektren hinterlassen können. So erwarten [Travis et al. \(2007\)](#) im Vergleich zur Ursprungspopulation mehr nachteilige Mutationen in neu besiedelten Bereichen. In echten genetischen Daten konnte eine solche Häufung nachteiliger Allele in Europa nachgewiesen werden ([Lohmueller et al., 2008](#)).

In meinen Analysen echter genetischer Daten habe ich mich auf Einzelnukleotid-Polymorphismen (SNPs) beschränkt, da hier eine große Anzahl an Datensätzen vorhanden war. Zukünftige Genomprojekte werden voraussichtlich vergleichbare Daten für unterschiedlich schnell mutierende Genorte (z.B. Mikrosatelliten) generieren. Dadurch wird eine genauere Unterscheidung von Ongoing-Mutations und Standing-genetic-Variation möglich werden ([Hofer et al., 2009](#)).

5. Literaturverzeichnis

- 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. *Nature*, 467(7319):1061–1073. [10](#), [29](#), [42](#), [60](#), [96](#)
- 1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T., and McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56–65. [10](#), [16](#), [59](#), [60](#), [62](#), [96](#), [98](#)
- Allee, W. C. (1931). *Animal Aggregations. A study in General Sociology*. The University of Chicago Press. [11](#)
- Allee, W. C. (1941). Integration of problems concerning protozoan populations with those of general biology. *The American Naturalist*, 75(760):pp. 473–487. [11](#), [25](#)
- Allendorf, F. W. and Lundquist, L. L. (2003). Introduction: Population biology, evolution, and control of invasive species. *Conservation Biology*, 17(1):24–30. [3](#)
- Anderson, A. and Hallatschek, O. (2014). Mutant size distributions in 2D Eden models. [113](#)
- Ardlie, K. G., Kruglyak, L., and Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet*, 3(4):299–309. [61](#)
- Arenas, M., Ray, N., Currat, M., and Excoffier, L. (2012). Consequences of range contractions and range shifts on molecular diversity. *Mol Biol Evol*, 29(1):207–218. [62](#)
- Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science*, 332(6027):346–349. [7](#)
- Austerlitz, F., Jung-Muller, B., Godelle, B., and Gouyon, P.-H. (1997). Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology*, 51(2):148 – 164. [7](#)
- Belle, E. M. S., Landry, P.-A., and Barbujani, G. (2006). Origins and evolution of the Europeans’ genome: evidence from multiple microsatellite loci. *Proc Biol Sci*, 273(1594):1595–1602. [6](#)

5. Literaturverzeichnis

- Biswas, S. and Akey, J. M. (2006). Genomic insights into positive selection. *Trends Genet*, 22(8):437–446. 8
- Bocxlaer, I. V., Loader, S. P., Roelants, K., Biju, S. D., Menegon, M., and Bossuyt, F. (2010). Gradual adaptation toward a range-expansion phenotype initiated the global radiation of toads. *Science*, 327(5966):679–682. 3
- Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H., and Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*, 140(2):783–796. 8
- Brockmann, D., Hufnagel, L., and Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075):462–465. 63
- Burton, O. J. and Travis, J. M. J. (2008). Landscape structure and boundary effects determine the fate of mutations occurring during range expansions. *Heredity (Edinb)*, 101(4):329–340. 13, 27, 44, 59
- Charlesworth, D. and Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, 10(11):783–796. 11
- Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B., and Thomas, C. D. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science*, 333(6045):1024–1026. 1, 2
- Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703. 54, 56
- Colautti, R. I. and Barrett, S. C. H. (2013). Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science*, 342(6156):364–366. 3
- Conrad, D. F., Keebler, J. E. M., DePristo, M. A., Lindsay, S. J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C. L., Torroja, C., Garimella, K. V., Zilversmit, M., Cartwright, R., Rouleau, G. A., Daly, M., Stone, E. A., Hurles, M. E., Awadalla, P., and , . G. P. (2011). Variation in genome-wide mutation rates within and between human families. *Nat Genet*, 43(7):712–714. 6, 10, 24, 29, 42, 59, 61
- Courchamp, Clutton-Brock, and Grenfell (1999). Inverse density dependence and the Allee effect. *Trends Ecol Evol*, 14(10):405–410. 11, 25, 84, 85
- Coventry, A., Bull-Otterson, L. M., Liu, X., Clark, A. G., Maxwell, T. J., Crosby, J., Hixson, J. E., Rea, T. J., Muzny, D. M., Lewis, L. R., Wheeler, D. A., Sabo, A., Lusk, C., Weiss, K. G., Akbar, H., Cree, A., Hawes, A. C., Newsham, I., Varghese, R. T.,

- Villasana, D., Gross, S., Joshi, V., Santibanez, J., Morgan, M., Chang, K., Iv, W. H., Templeton, A. R., Boerwinkle, E., Gibbs, R., and Sing, C. F. (2010). Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat Commun*, 1:131. 8, 22, 29
- Currat, M. and Excoffier, L. (2005). The effect of the Neolithic expansion on European molecular diversity. *Proc Biol Sci*, 272(1564):679–688. 6, 17, 22, 27, 30, 59, 79
- Currat, M., Excoffier, L., Maddison, W., Otto, S. P., Ray, N., Whitlock, M. C., and Yeaman, S. (2006). Comment on “Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*” and “Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans”. *Science*, 313(5784):172; author reply 172. 1, 8
- Davis, M. A., Chew, M. K., Hobbs, R. J., Lugo, A. E., Ewel, J. J., Vermeij, G. J., Brown, J. H., Rosenzweig, M. L., Gardener, M. R., Carroll, S. P., Thompson, K., Pickett, S. T. A., Stromberg, J. C., Tredici, P. D., Suding, K. N., Ehrenfeld, J. G., Grime, J. P., Mascaro, J., and Briggs, J. C. (2011). Don’t judge species on their origins. *Nature*, 474(7350):153–154. 1, 3
- Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer. 22, 27, 88
- Durrett, R. (2013). Population genetics of neutral mutations in exponentially growing cancer cell populations. *Ann Appl Probab*, 23(1):230–250. 8, 34, 55, 113
- Dyer, D. W. (2010). Uncommons maths. <http://maths.uncommons.org/>. 16, 88
- Easteal, S. (1981). The history of introductions of *Bufo marinus* (Amphibia: Anura); a natural experiment in evolution. *Biological Journal of the Linnean Society*, 16(2):93–113. 4
- Edmonds, C. A., Lillie, A. S., and Cavalli-Sforza, L. L. (2004). Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America*, 101(4):975–979. 7, 8, 10, 17, 22, 27, 33, 44, 53, 54
- Eswaran, V. (2002). A diffusion wave out of Africa: The mechanism of the modern human revolution? *Current Anthropology*, 43(5):pp. 749–774. 25
- Excoffier, L., Foll, M., and Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):481–501. 1, 7, 8
- Excoffier, L. and Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol*, 23(7):347–351. 1, 9

5. Literaturverzeichnis

- Fayard, J., Klein, E. K., and Lefèvre, F. (2009). Long distance dispersal and the fate of a gene from the colonization front. *J Evol Biol*, 22(11):2171–2182. 62
- Gralka, M., Anderson, A., and Hallatschek, O. (2014). Unpublished results. 113
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspina, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L. F., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., and Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, 328(5979):710–722. 5
- Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K. W., and Burger, J. (2005). Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science*, 310(5750):1016–1018. 6
- Hallatschek, O., Hersen, P., Ramanathan, S., and Nelson, D. R. (2007). Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50):19926–19930. 2, 4, 8, 54, 116
- Hallatschek, O. and Nelson, D. R. (2008). Gene surfing in expanding populations. *Theor Popul Biol*, 73(1):158–170. 9
- Hartl, D. L., Clark, A. G., et al. (2007). *Principles of population genetics*, volume 116. Sinauer associates, Sunderland, Massachusetts, fourth edition. 15, 18, 19, 34, 56, 57, 58, 75
- Hassold, T., Sherman, S., and Hunt, P. (2000). Counting cross-overs: characterizing meiotic recombination in mammals. *Hum Mol Genet*, 9(16):2409–2419. 61
- Hebert, P. D. N., Muncaster, B. W., and Mackie, G. L. (1989). Ecological and genetic studies on *Dreissena polymorpha* (Pallas): a new mollusc in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 46(9):1587–1591. 2, 4
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789):907–913. 1, 2, 4

- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58(3):247–276. [2](#), [4](#)
- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc Lond B Biol Sci*, 359(1442):183–95; discussion 195. [2](#)
- Hofer, T., Ray, N., Wegmann, D., and Excoffier, L. (2009). Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Ann Hum Genet*, 73(1):95–108. [1](#), [8](#), [63](#)
- Holsinger, K. E. and Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting f_{st} . *Nature Reviews Genetics*, 10(9):639–650. [11](#), [35](#)
- Kampen, N. V. (2007). *Stochastic Processes in Physics and Chemistry, Third Edition (North-Holland Personal Library)*. North Holland. [115](#), [116](#)
- Keinan, A. and Clark, A. G. (2012). Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*, 336(6082):740–743. [8](#), [10](#), [55](#)
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626. [63](#)
- Kimura, M. and Ohta, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61(3):763–771. [56](#), [57](#)
- Klopfstein, S., Currat, M., and Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, 23(3):482–490. [7](#), [10](#), [17](#), [22](#), [27](#), [30](#), [33](#), [44](#), [49](#), [56](#), [57](#), [58](#), [79](#)
- Krehenwinkel, H. and Tautz, D. (2013). Northern range expansion of European populations of the wasp spider *Argiope bruennichi* is associated with global warming-correlated genetic admixture and population-specific temperature adaptations. *Mol Ecol*, 22(8):2232–2248. [3](#)
- Kubisch, A., Holt, R. D., Poethke, H.-J., and Fronhofer, E. A. (2013). Where am I and why? Synthesizing range biology and the eco-evolutionary dynamics of dispersal. *Oikos*, no:no–no. [2](#)
- Kubisch, A., Poethke, H.-J., and Hovestadt, T. (2011). Density-dependent dispersal and the formation of range borders. *Ecography*, 34(6):1002–1008. [84](#), [85](#)
- Langergraber, K. E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E., Inoue-Muruyama, M., Mitani, J. C., Muller, M. N., Robbins, M. M., Schubert,

5. Literaturverzeichnis

- G., Stoinski, T. S., Viola, B., Watts, D., Wittig, R. M., Wrangham, R. W., Zuberbühler, K., Pääbo, S., and Vigilant, L. (2012). Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc Natl Acad Sci U S A*, 109(39):15716–15721. [31](#), [78](#)
- Lev-Yadun, S., Gopher, A., and Abbo, S. (2000). Archaeology. the cradle of agriculture. *Science*, 288(5471):1602–1603. [6](#)
- Liebhold, A. and Bascompte, J. (2003). The Allee effect, stochastic dynamics and the eradication of alien species. *Ecology Letters*, 6(2):133–140. [85](#)
- Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S. W., Yamaoka, Y., Graham, D. Y., Perez-Trallero, E., Wadstrom, T., Suerbaum, S., and Achtman, M. (2007). An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*, 445(7130):915–918. [1](#), [7](#)
- Lodish, H., Berk, A., Matsudaira, P., Kaiser, C. A., Krieger, M., Scott, M. P., Zipursky, L., and Darnell, J. (2003). *Molecular Cell Biology*. W. H. Freeman. [61](#)
- Lohmueller, K. E., Indap, A. R., Schmidt, S., Boyko, A. R., Hernandez, R. D., Hubisz, M. J., Sninsky, J. J., White, T. J., Sunyaev, S. R., Nielsen, R., Clark, A. G., and Bustamante, C. D. (2008). Proportionally more deleterious genetic variation in European than in African populations. *Nature*, 451(7181):994–997. [63](#)
- Manica, A., Amos, W., Balloux, F., and Hanihara, T. (2007). The effect of ancient population bottlenecks on human phenotypic variation. *Nature*, 448(7151):346–348. [7](#)
- Matsumoto, M. and Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.*, 8(1):3–30. [16](#)
- Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., and Worm, B. (2011). How many species are there on earth and in the ocean? *PLoS Biol*, 9(8):e1001127. [1](#)
- Neher, R. A. and Hallatschek, O. (2013). Genealogies of rapidly adapting populations. *Proc Natl Acad Sci U S A*, 110(2):437–442. [55](#)
- Newman, M. (2005). Power laws, Pareto distributions and Zipf’s law. *Contemporary Physics*, 46(5):323–351. [31](#), [32](#), [45](#), [54](#), [90](#), [94](#), [96](#)
- Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A. G. (2007). Recent and ongoing selection in the human genome. *Nat Rev Genet*, 8(11):857–868. [1](#), [8](#), [10](#), [13](#), [53](#)

- Nowak, M. A. (2006). *Evolutionary Dynamics: Exploring the Equations of Life*. Belknap Press. 21, 77
- Nullmeier, J. and Hallatschek, O. (2013). The coalescent in boundary-limited range expansions. *Evolution*, 67(5):1307–1320. 56, 116
- Olivieri, I. (2009). Alternative mechanisms of range expansion are associated with different changes of evolutionary potential. *Trends Ecol Evol*, 24(6):289–292. 1, 11
- Oppenheimer, S. (2012). Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. *Philos Trans R Soc Lond B Biol Sci*, 367(1590):770–784. 2, 3, 4, 5, 59
- Oracle Corporation (2011). Java SE 6 documentation. <http://docs.oracle.com/javase/6/docs/>. 16
- Phillips, B. L. and Shine, R. (2005). The morphology, and hence impact, of an invasive species (the cane toad, *Bufo marinus*): changes with time since colonisation. *Animal Conservation*, 8(4):407–413. 2
- Pimentel, D., Zuniga, R., and Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, 52(3):273 – 288. Integrating Ecology and Economics in Control Bioinvasions {IEECB} S.I. 3
- Pinhasi, R., Thomas, M. G., Hofreiter, M., Currat, M., and Burger, J. (2012). The genetic history of Europeans. *Trends Genet*, 28(10):496–505. 2, 6
- Prugnolle, F., Manica, A., and Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. *Curr Biol*, 15(5):R159–R160. 7, 25
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 16
- Rasmussen, M., Anzick, S. L., Waters, M. R., Skoglund, P., DeGiorgio, M., Stafford Jr, T. W., Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S. M., Poznik, G. D., Gudmundsdottir, V., Yadav, R., Malaspinas, A.-S., V, S. S. W., Allentoft, M. E., Cornejo, O. E., Tambets, K., Eriksson, A., Heintzman, P. D., Karmin, M., Korneliussen, T. S., Meltzer, D. J., Pierre, T. L., Stenderup, J., Saag, L., Warmuth, V. M., Lopes, M. C., Malhi, R. S., Brunak, S., Sicheritz-Ponten, T., Barnes, I., Collins, M., Orlando, L., Balloux, F., Manica, A., Gupta, R., Metspalu, M., Bustamante, C. D., Jakobsson, M., Nielsen, R., and Willerslev, E. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature*, 506(7487):225–229. 5

5. Literaturverzeichnis

- Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., Gilbert, M. T. P., Wang, Y., Raghavan, M., Campos, P. F., Kamp, H. M., Wilson, A. S., Gledhill, A., Tridico, S., Bunce, M., Lorenzen, E. D., Binladen, J., Guo, X., Zhao, J., Zhang, X., Zhang, H., Li, Z., Chen, M., Orlando, L., Kristiansen, K., Bak, M., Tommerup, N., Bendixen, C., Pierre, T. L., Grønnow, B., Meldgaard, M., Andreasen, C., Fedorova, S. A., Osipova, L. P., Higham, T. F. G., Ramsey, C. B., Hansen, T. V. O., Nielsen, F. C., Crawford, M. H., Brunak, S., Sicheritz-Pontén, T., Villems, R., Nielsen, R., Krogh, A., Wang, J., and Willerslev, E. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, 463(7282):757–762. [6](#)
- Ray, N. and Excoffier, L. (2010). A first step towards inferring levels of long-distance dispersal during past expansions. *Mol Ecol Resour*, 10(5):902–914. [22](#), [27](#)
- Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M. V., Derevianko, A. P., Hublin, J.-J., Kelso, J., Slatkin, M., and Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468(7327):1053–1060. [1](#), [5](#)
- Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., Ko, A. M.-S., Ko, Y.-C., Jinam, T. A., Phipps, M. E., Saitou, N., Wollstein, A., Kayser, M., Pääbo, S., and Stoneking, M. (2011). Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am J Hum Genet*, 89(4):516–528. [5](#)
- Scally, A. and Durbin, R. (2012). Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet*, 13(10):745–753. [6](#)
- Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A. S., Cavalli-Sforza, L. L., and Underhill, P. A. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science*, 290(5494):1155–1159. [7](#)
- Sibly, R. M., Barker, D., Denham, M. C., Hone, J., and Pagel, M. (2005). On the regulation of populations of mammals, birds, fish, and insects. *Science*, 309(5734):607–610. [84](#)
- Slatkin, M. and Excoffier, L. (2012). Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics*, 191(1):171–181. [7](#)

- Stoneking, M. and Krause, J. (2011). Learning about human population history from ancient and modern genomes. *Nat Rev Genet*, 12(9):603–614. 2, 3
- Taylor, C. M. and Hastings, A. (2005). Allee effects in biological invasions. *Ecology Letters*, 8(8):895–908. 25, 57, 85
- Taylor, P. D. and Jonker, L. B. (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1–2):145 – 156. 77
- Thompson, D. Q., Stuckey, R. L., and Thompson, E. B. (1987). Spread, impact, and control of purple loosestrife (*Lythrum salicaria*) in North American wetlands. 55 pages. Jamestown, ND: Northern Prairie Wildlife Research Center Online. <http://www.npwrc.usgs.gov/resource/plants/loosstrf/index.htm> (Version 04JUN1999). 2, 3, 4
- Travis, J. M. J., Münkemüller, T., Burton, O. J., Best, A., Dytham, C., and Johst, K. (2007). Deleterious mutations can surf to high densities on the wave front of an expanding population. *Molecular Biology and Evolution*, 24(10):2334–2343. 8, 10, 17, 33, 44, 58, 63
- Verhulst, P. F. (1838). Notice sur la loi que la population poursuit dans son accroissement. *Correspondance mathématique et physique*, 10:113–121. 11, 82
- Verhulst, P.-F. (1845). Recherches mathématiques sur la loi d’accroissement de la population. *Nouveaux Mémoires de l’Académie Royale des Sciences et Belles-Lettres de Bruxelles*, 18:1–42. Mathematical Researches into the Law of Population Growth Increase. 11, 82
- Wacey, D., Kilburn, M. R., Saunders, M., Cliff, J., and Brasier, M. D. (2011). Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nature Geoscience*, 4(10):698–702. 1
- Wakeley, J. (2008). *Coalescent Theory: An Introduction*. Roberts and Company Publishers. 34, 55, 113
- Wang, G., Liang, X.-G., and Wang, F.-Z. (1999). The competitive dynamics of populations subject to an Allee effect. *Ecological Modelling*, 124(2–3):183 – 192. 85
- Waters, J. M., Fraser, C. I., and Hewitt, G. M. (2013). Founder takes all: density-dependent processes structure biodiversity. *Trends Ecol Evol*, 28(2):78–85. 63
- Wegmann, D., Currat, M., and Excoffier, L. (2006). Molecular diversity after a range expansion in heterogeneous environments. *Genetics*, 174(4):2009–2020. 22, 27, 60

5. Literaturverzeichnis

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16(2):97–159. 15, 18, 19

A. Anhang (Methoden)

Dieses Kapitel enthält zusätzliche Informationen und Beispiele zum verwendeten Wright-Fisher Modell (ABSCHN. A.1), Hintergründe und Quelltexte zur Implementierung des Modells (ABSCHN. A.2) sowie Details zur Auswertung (ABSCHN. A.3).

A.1. Details zum Modell

Die Dynamik innerhalb der Deme habe ich mit Hilfe des Wright-Fisher Modells simuliert. Im Wright-Fisher Modell ist die Populationsgröße auf K Individuen begrenzt, ansonsten sind alle Voraussetzungen einer Population des Hardy-Weinberg-Modells erfüllt (Hartl et al., 2007). Durch diese Begrenzung der Populationsgröße kann sich die Zusammensetzung der Population von Generation zu Generation durch genetische Drift ändern.

Eine Annahme im Wright-Fisher Modell ist, dass es keine Koexistenz der Eltern- und der Nachkommengeneration gibt. Das trifft z.B. auf einjährige Pflanzen zu. Um die Zusammensetzung der Population in der nächsten Generation zu bestimmen, wird jede Kopie eines Gens in der neuen Generation unabhängig und zufällig aus dem Genpool der vorherigen Generation gezogen (Ziehen aus einer Binomialverteilung). In ABB. A.1 ist dieser Ablauf für haploide Individuen gezeigt, die am betrachteten Genort entweder ein „orangefarbenes Allel“ oder ein „blaues Allel“ haben. Die Modellierung von Populationen mit K diploiden Individuen entspricht im Wesentlichen einem Modell mit $2K$ haploiden Individuen (Hartl et al., 2007).

Bei einer Range-Expansion wird ein Gebiet schrittweise besiedelt. Daher habe ich aus dem Wright-Fisher Modell weitere Modelle abgeleitet, in denen die Deme auf einem zweidimensionalen Gitter angeordnet sind (ABB. 2.1). Ausgehend von einer Ursprungspopulation werden die Deme schrittweise besiedelt (ABB. 2.3, ABB. 2.4).

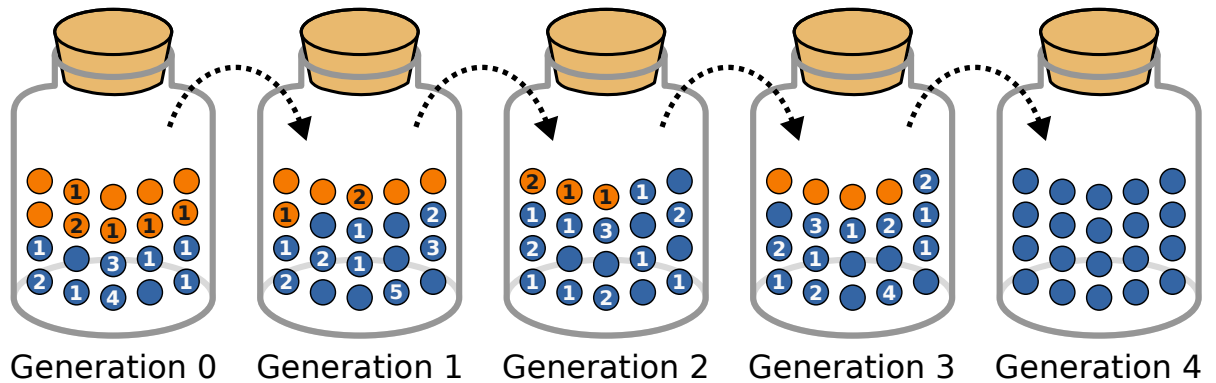


Abbildung A.1.: **Wright-Fisher Modell für zufällige genetische Drift.** Dargestellt ist die Ermittlung der Populationszusammensetzung über fünf Generationen. Im Beispiel wird jede Generation durch ein Glas symbolisiert, in dem sich eine Population von $K = 20$ Individuen befindet. Diese Individuen unterscheiden sich in ihrem Genotyp und setzen sich zu Beginn aus 50% blauen und 50% orangefarbenen Individuen zusammen. Zur Ermittlung der nächsten Generation wird K -mal eine Kugel aus der aktuellen Generation gezogen, dann eine Kugel des gleichen Genotyps in das neue Glas gelegt und im Anschluss die gezogene Kugel zurückgelegt. Im obigen Beispiel gibt die Zahl innerhalb der Kugel an, wie oft diese Kugel gezogen und wieder zurückgelegt wurde, also wie viele Nachkommen sie in der nächsten Generation hat. Dieser Vorgang entspricht binomialem Sampling („Ziehen mit Zurücklegen“). Allein durch diese Zufallsereignisse ändert sich die Zusammensetzung des Genpools. Ist ein Genotyp in allen Individuen vorhanden, so spricht man von der *Fixierung* des Allels (hier: das blaue Allel in Generation 4). Sofern keine Mutationen auftreten, setzen sich alle Folgegenerationen nur noch aus Individuen mit dem blauen Allel zusammen. Bildvorlage: KAP. C.

Beispiel-Simulationen Für die Simulationen habe ich unterschiedliche Gitter verwendet, die in TAB. 2.2 zusammengefasst sind. Zur Illustration habe ich Momentaufnahmen einzelner Simulationen auf der „Kunstwelt“ (ABB. A.2), der „Europakarte“ (ABB. A.3) und der „Großen Welt“ (ABB. A.4) zu verschiedenen Zeitpunkten erstellt.

A.2. Implementierungen

A.2.1. Implementierung von Selektion

In der Regel habe ich neutrale Mutationen simuliert, das Wachstum konnte ich hingegen mit Hilfe von Selektion realisieren (siehe TEILABSCHN. 2.1.2). Wie in GL. 2.2 beschrieben entsprechen die erwarteten Frequenzen neutraler Mutationen in der Nachkommengeneration den Frequenzen der aktuellen Generation. Spielt allerdings Selektion (bzw. Wachstum) eine Rolle, so ändert sich die erwartete Frequenz p'_i einer Variante i im Vergleich zur Frequenz p_i der aktuellen Generation um einen Wert Δp_i :

$$p'_i = p_i + \Delta p_i. \quad (\text{A.1})$$

Für die Berechnung von Δp_i habe ich die diskrete Form der Replikator-Gleichung verwendet (Nowak, 2006; Taylor and Jonker, 1978). Die Replikator-Gleichung beschreibt die Dynamik einer Population, in der Individuen n verschiedener Typen in Konkurrenz zueinander stehen. Die Änderung Δp_i der Frequenz p_i der Individuen des Typs i ist gegeben durch:

$$\Delta p_i = p_i [s_i - \phi] \text{ für } i = 1, \dots, n \quad (\text{A.2})$$

mit

$$\sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n \Delta p_i = 0.$$

Dabei entspricht ϕ der durchschnittlichen Fitness der Population, welche durch die gewichteten Fitness-Werte s_i der einzelnen Typen errechnet wird:

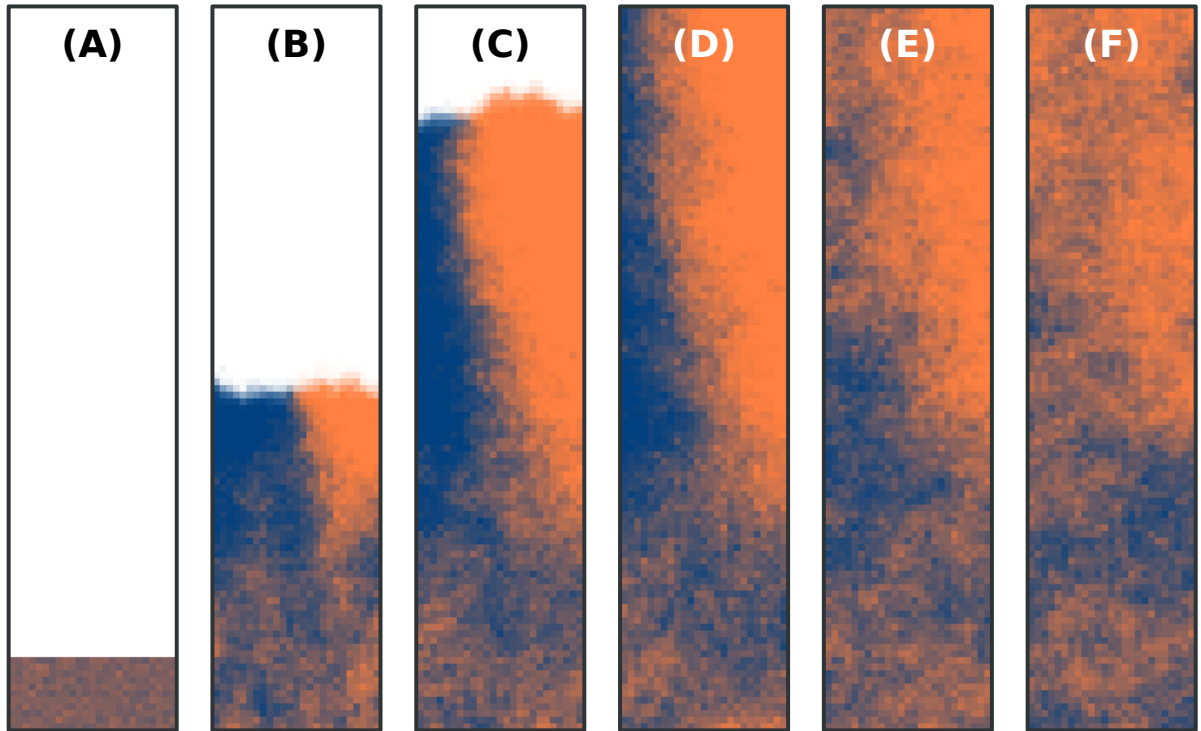


Abbildung A.2.: **Momentaufnahmen einer Range-Expansion Simulation auf der „Kunstwelt“**. Die Bilder zeigen die Verteilung der Allele zu Beginn der Simulation (A), nach 150 (B), 300 (C), 450 (D), 1000 (E) und 2000 (F) Generationen. Die Ausmaße des Gitters waren $w = 25$ mal $h = 109$ Deme, wobei zu Beginn der Simulation die ersten 11 Reihen besiedelt waren. In diesem Beispiel besteht die Anfangspopulation zu 50% ($p_s = 0.5$) aus Individuen mit dem abgeleiteten Allel (orange). Jedes Dem bot Platz für maximal $K = 100$ Individuen, wobei das Wachstum innerhalb der Deme logistisch reguliert war (Wachstumsrate: $r = 0.5$). In jeder Generation sind 20% ($m = 0.2$) der Individuen eines Dems mit benachbarten Demen ausgetauscht worden. Die Simulationen liefen für $g = 2000$ Generationen, was einer Zeitspanne von 50'000 – 60'000 Jahren für den Menschen entspricht (Generationszeit: 25 – 30 Jahre; [Langergraber et al. \(2012\)](#)).

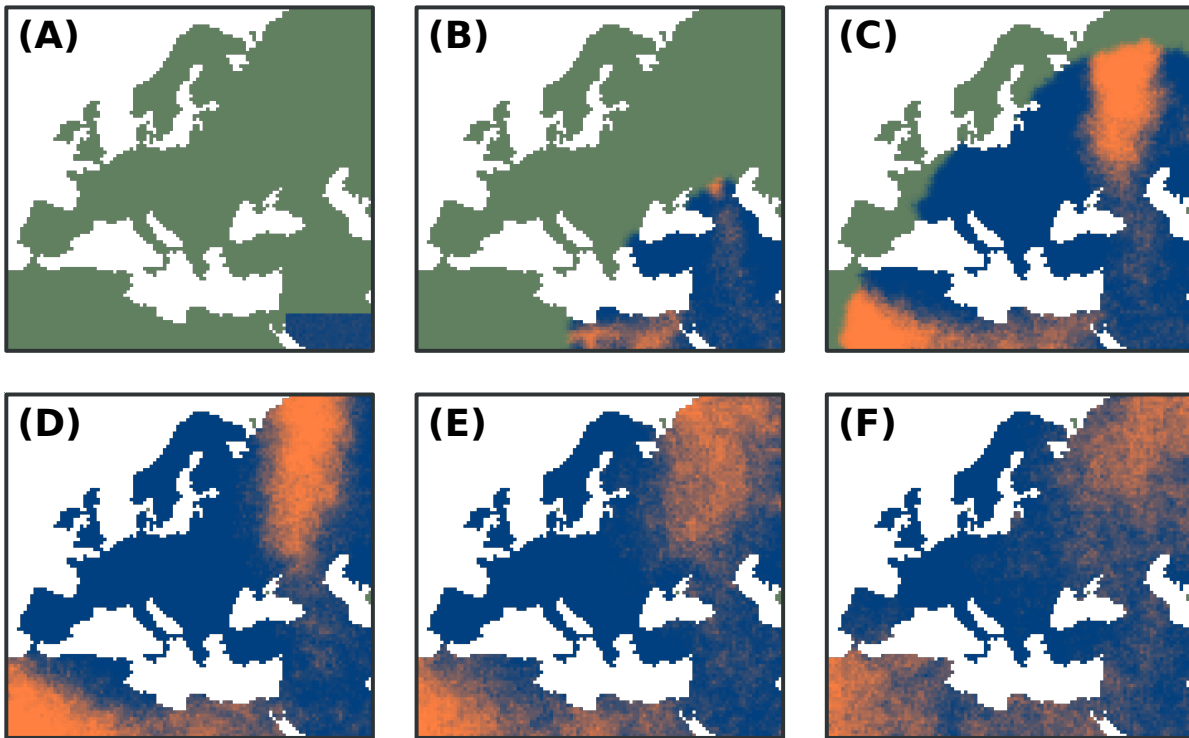


Abbildung A.3.: Momentaufnahmen einer Range-Expansion Simulation auf der „Europakarte“. Gezeigt ist die Verteilung des ursprünglichen (blau) und des abgeleiteten Allels (orange) zu Beginn der Simulation (A), nach 150 (B), 300 (C), 450 (D), 1000 (E) und 2000 (F) Generationen. Die Landmasse war in 7223 bewohnbare Deme unterteilt, wodurch jedes Dem in etwa eine Fläche von 2500km^2 repräsentiert (Klopfstein et al., 2006; Currat and Excoffier, 2005). Die Ursprungspopulation befand sich im Nahen Osten, einem Bereich, von dem aus der Mensch vermutlich den Europäischen Kontinent besiedelt hat (Currat and Excoffier, 2005). In diesem Beispiel waren von den 27'000 Individuen der Ausgangspopulation 10% ($p_s = 0.1$) Träger des abgeleiteten Allels (orange). Die übrigen Parameter entsprechen denen aus ABB. A.2: $K = 100$; $r = 0.5$; $m = 0.2$ und $g = 2000$.

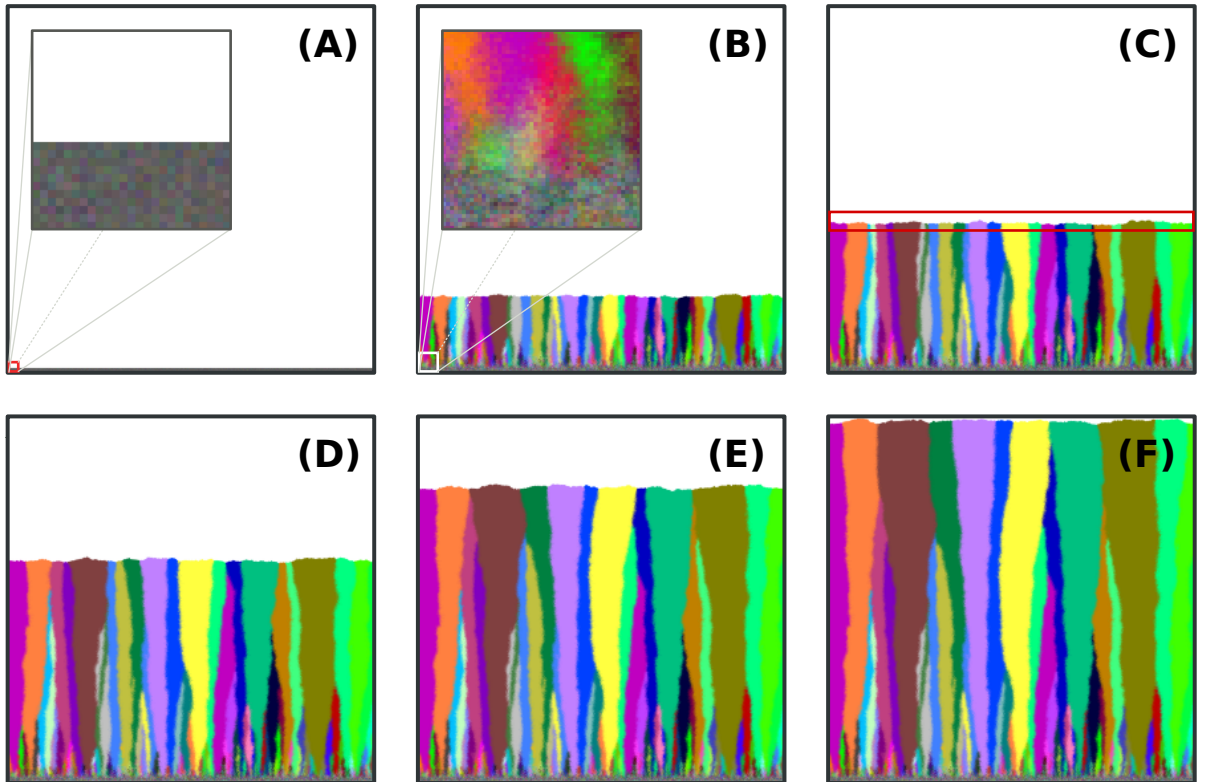


Abbildung A.4.: **Momentaufnahmen einer Range-Expansion Simulation auf der „Großen Welt“** (Breite $w = 1000$, Höhe $h = 1000$). In diesem Beispiel konnte jedes Dem $K = 100$ Individuen enthalten, wodurch auf diesem Gitter insgesamt 100 Millionen Individuen Platz hatten. Zu Beginn der Simulation waren die ersten 11 Reihen mit Individuen aus 124 verschiedenen genetischen Varianten besiedelt, deren Anfangsfrequenz jeweils 0.807% ($p_s = 1/124$) war. Dargestellt ist die Verteilung dieser Varianten zu Beginn der Simulation (A), nach 700 (B), 1400 (C), 2100 (D), 2800 (E) Generationen und als ein Individuum die letzte Reihe erreicht hatte (F; hier nach 3432 Generationen). In (A) zeigt das eingefügte Bild einen vergrößerten Ausschnitt von 25×25 Demen. In (B) zeigt das eingefügte Bild einen vergrößerten Ausschnitt von 50×50 Demen. Im Gegensatz zu **ABB. A.2** und **ABB. A.3** fand die Migration auf diesem Gitter nur in einem begrenzten Bereich statt, der sich mit der Front bewegte (roter Kasten in C). Daher spielte Drift in den Bereichen hinter der Front keine Rolle und es fand kein Austausch von Individuen zwischen einzelnen Demen statt. ($m = 0.2$; $r = 0.5$; $\alpha = n.a.$; $\epsilon = n.a.$).

$$\phi = \sum_{i=1}^n p_i s_i. \quad (\text{A.3})$$

Setzt man GL. A.2 und GL. A.3 in GL. A.1 ein, so kann man die erwartete Frequenz p'_i einer Variante i in der nächsten Generation errechnen:

$$\begin{aligned} p'_i &= p_i + p_i \left[s_i - \sum_{j=1}^n p_j s_j \right] \\ &= p_i \left[1 + s_i - \sum_{i=1}^n p_i s_i \right]. \end{aligned}$$

Bei $n = 3$ Typen ergibt sich die Frequenz des Typen $i = 1$ in der nächsten Generation aus:

$$\begin{aligned} p'_1 &= p_1 \left[1 + \underbrace{s_1 - p_1 s_1}_{=s_1(1-p_1)} - p_2 s_2 - p_3 s_3 \right] \\ &= p_1 \left[1 + \underbrace{s_1(1-p_1)}_{=p_2+p_3} - p_2 s_2 - p_3 s_3 \right] \\ &= p_1 \left[1 + s_1 p_2 + s_1 p_3 - p_2 s_2 - p_3 s_3 \right] \\ &= p_1 \left[1 + p_2 \underbrace{(s_1 - s_2)}_{\equiv s_{1,2}} + p_3 \underbrace{(s_1 - s_3)}_{\equiv s_{1,3}} \right] \\ &= p_1 + p_1 p_2 s_{1,2} + p_1 p_3 s_{1,3}. \end{aligned} \quad (\text{A.4})$$

Die Differenz der beiden Fitnesswerte s_i und s_j wird ausgedrückt durch $s_{i,j}$, mit $s_{i,j} = -s_{j,i}$. GL. A.4 lässt sich allgemein formulieren als:

$$p'_i = p_i + \sum_{i=0}^{n-1} p_i p_j s_{i,j}. \quad (\text{A.5})$$

Beim Vergleich von GL. A.5 mit GL. A.9 aus TEILABSCHN. A.2.2 ist zu erkennen, dass in einer Population mit Individuen der Varianten i und j der Anteil der Individuen der Variante i logistisch gegenüber dem Anteil der Variante j mit einer Wachstumsrate

A. Anhang (Methoden)

$s_{i,j}$ wächst bzw. schrumpft.

A.2.2. Implementierung des Populationswachstums

Als mathematisches Modell für das Wachstum der Population habe ich das logistische Wachstumsmodell nach Pierre-François Verhulst verwendet (Verhulst, 1838, 1845). Die Gleichung GL. A.6 gibt die aktuelle Anzahl $N(t)$ der Individuen einer Population zum Zeitpunkt t wieder. Diese ist abhängig von der anfänglichen Anzahl N_0 der Individuen in der Population, der Kapazitätsgrenze K sowie einer Wachstumsrate r :

$$N(t) = \frac{1}{\left(\frac{1}{N_0} - \frac{1}{K}\right) e^{-rt} + \frac{1}{K}}. \quad (\text{A.6})$$

Die Änderung $dN(t)$ der Individuenzahl über eine Zeitdifferenz dt wird beschrieben durch die Zeitableitung der Verhulst-Gleichung:

$$\begin{aligned} \frac{dN(t)}{dt} &= rN(t) - \frac{r}{K} \cdot N(t)^2 \\ &= \underbrace{rN(t) \cdot \left(1 - \frac{N(t)}{K}\right)}_{f(N)}. \end{aligned} \quad (\text{A.7})$$

Bei Computersimulationen wird immer in diskreten Zeitschritten gerechnet. Im Folgenden entspricht N der Zahl der Individuen zum Zeitpunkt t und N' der Zahl der Individuen im nächsten simulierten Zeitschritt ($t + 1$). Die Differenz zum nächsten Zeitschritt ist genau 1.

$$\begin{aligned} dt &\mapsto \Delta t = 1 \\ dN(t) &\mapsto \Delta N = N' - N \end{aligned}$$

Für die diskretisierte Form der Zeitableitung der Verhulst-Gleichung ergibt sich also:

$$\begin{aligned}
\frac{N' - N}{\Delta t} &= f(N) \\
N' &= N + \underbrace{\Delta t}_{=1} \cdot f(N) \\
\frac{N'}{K} &= \frac{N}{K} + \frac{f(N)}{K} \\
p' &= p + \underbrace{\frac{1}{K} \cdot f(N)}_{\text{Wachstum}}.
\end{aligned} \tag{A.8}$$

Die Populationsdichte p' im nächsten Zeitschritt errechnet sich also aus der aktuellen Populationsdichte p plus der Änderung durch das Wachstum. Setzt man nun $f(N)$ aus GL. A.7 in GL. A.8 ein, so erhält man:

$$\begin{aligned}
p' &= p + \frac{1}{K} \cdot \left(rN \cdot \left(1 - \frac{N}{K} \right) \right) \\
&= p + p \cdot \underbrace{r \left(1 - p \right)}_{\text{Pro-Kopf-Zuwachs}}.
\end{aligned} \tag{A.9}$$

Nun vergleiche ich das logistische Wachstum aus GL. A.9 mit der Änderung der Allelfrequenzen aus der Replikator-Gleichung (GL. A.5). Fasse ich dabei den noch nicht besiedelten Anteil des Doms ($1 - p$) als einem eigenen Typen auf, so entspricht die Wachstumsrate r genau dem Selektionskoeffizienten $s_{i,j}$ aus der Replikatorgleichung. Daher habe ich in meinen Simulationen das Wachstum über einen selektiven Nachteil der *leeren Felder* gegenüber den *genetischen Varianten* (bzw. *Individuen*) ausgedrückt (siehe auch TEILABSCHN. 2.1.2).

A.2.3. Implementierung des Allee Effektes

In GL. A.7 wird für das logistische Wachstum die Änderung der Individuenzahlen pro Zeiteinheit in der Gesamtpopulation betrachtet. Teilt man diesen Wert durch die Zahl der Individuen $N(t)$, die sich zu diesem Zeitpunkt in der Population befinden, so bekommt man die Pro-Kopf-Zuwachsrate $L(N)$:

A. Anhang (Methoden)

Tabelle A.1.: Liste der Pro-Kopf-Zuwachsraten verschiedener negativ-dichteabhängiger Wachstumsmodelle. Die Funktionen sind in [ABB. A.5](#) grafisch dargestellt. (r : Wachstumsrate; θ : Formparameter; $N(t)$: Zahl der Individuen zum Zeitpunkt t ; K : Kapazitätsgrenze)

Bezeichnung	Funktion	Abkürzung	Referenz
Logistisch	$r \cdot \left(1 - \frac{N(t)}{K}\right)$	$L(N)$	Courchamp et al. (1999)
θ -Logistisch	$r \cdot \left(1 - \left(\frac{N(t)}{K}\right)^\theta\right)$	$T(N)$	Sibly et al. (2005)
Beverton-Holt	$(1 + r) \cdot \frac{1}{1 + r \frac{N(t)}{K}}$	$B(N)$	Kubisch et al. (2011)

$$L(N) = \frac{dN(t)}{N(t) \cdot dt} = r \cdot \underbrace{\left(1 - \frac{N(t)}{K}\right)}_{L(N)}.$$

Beim logistischen Wachstum ist die Pro-Kopf-Zuwachsrates am größten, wenn die Populationsdichte $p = N(t)/K$ nahe Null ist. In diesem Falle hat jedes Individuum im Mittel etwa $1+r$ Nachkommen, da $1 - N(t)/K \approx 1$. Das logistische Wachstum ist somit ein *negativ-dichteabhängiges Wachstum*, da mit zunehmender Populationsdichte auch das Pro-Kopf Wachstum geringer wird. Weitere negativ-dichteabhängige Wachstumsfunktionen sind in [TAB. A.1](#) aufgelistet und in [ABB. A.5](#) grafisch dargestellt.

Beim Allee-Effekt handelt es sich um ein *positiv-dichteabhängiges Wachstum*, da die Wachstumsrate mit zunehmender Populationsdichte steigt. Dazu wird ein dichteabhängiger Faktor AF mit der Pro-Kopf-Zuwachsrates multipliziert, der für ein geringeres Wachstum bei einer niedrigen Populationsdichte sorgt. Diverse Implementierungen dieses Effektes sind in [TAB. A.2](#) aufgelistet und in [ABB. A.5](#) grafisch dargestellt.

Um den Allee-Effekt zu simulieren, habe ich den Faktor AF so gewählt, dass das Wachstum erst bei einer Populationsdichte α stattfand. War die Populationsdichte $p < \alpha$, so fand kein bzw. ein verringertes Wachstum statt. Bei einer Populationsdichte $p > \alpha$ wuchs die Population mit der logistischen Pro-Kopf-Wachstumsrate $L(N)$ (siehe [TAB. A.1](#)). Die Breite des Übergangsbereiches der Wachstumsrate von 0 zu $L(N)$ wurde mit Hilfe des Exponenten ϵ reguliert. In [ABB. A.6](#) ist der Faktor AF für verschiedene α und ϵ in Abhängigkeit von der Populationsdichte p dargestellt. Bei meiner Umsetzung handelt es sich um einen nicht-kritischen Allee-Effekt, da es kein negatives Wachstum

Tabelle A.2.: Liste einiger in der Literatur verwendeter Modelle für den Allee-Effekt. Die Funktionen sind in ABB. A.5 grafisch dargestellt. K_c : Kritische Kapazitätsgrenze (*critical carrying capacity*); C : Allee-Effekt Konstante (*Allee effect constant*); α : Populationsdichte, bei der das Pro-Kopf-Wachstum halb-maximal ist; ϵ : Parameter, der die Steilheit des Allee-Effektes bestimmt (Für α und ϵ siehe ABB. A.6); N : aktuelle Größe der Population; K : Kapazitätsgrenze. Die Funktionen $L(N)$ und $B(N)$ sind in TAB. A.1 aufgeführt.

Bezeichnung	Funktion	Referenz
Allee	$L(N) \cdot \left(\frac{N}{K_c} - 1\right)$	Courchamp et al. (1999)
Wang	$L(N) \cdot \left(\frac{N}{N+C}\right)$	Wang et al. (1999)
Kubisch	$\left[B(N) \cdot \left(\frac{(N/K)^2}{(N/K)^2 + \alpha^2}\right)\right] - 1$	Kubisch et al. (2011)
Dennis	$L(N) \cdot \left(\frac{N-C}{K}\right)$	Liebhold and Bascompte (2003)
Boekhoff	$L(N) \cdot \left(\frac{(N/K)^\epsilon}{(N/K)^\epsilon + \alpha^\epsilon}\right)$	TEILABSCHN. 2.1.2

für geringe Populationsdichten gibt ([Taylor and Hastings, 2005](#)).

A.2.4. Implementierung der Zufallsverteilungen

Um die Zusammensetzung der Population in der nächsten Generation zu bestimmen, habe ich aus einer Multinomialverteilung gezogen („Ziehen mit Zurücklegen“; siehe CODE A.1). Die dazu benötigten erwarteten Wahrscheinlichkeiten habe ich zuvor aus den aktuellen Allelfrequenzen, der Wachstumsrate sowie der Stärke des Allee-Effektes errechnet (siehe GL. 2.3 und GL. 2.5).

Die Migranten habe ich multi-hypergeometrisch gezogen („Ziehen ohne Zurücklegen“; siehe CODE A.2). Damit konnte ich sicherstellen, dass nicht zwei Individuen eines Typs zur Migration ausgewählt werden, obwohl nur ein Individuum dieses Typs in der Population vorhanden ist.

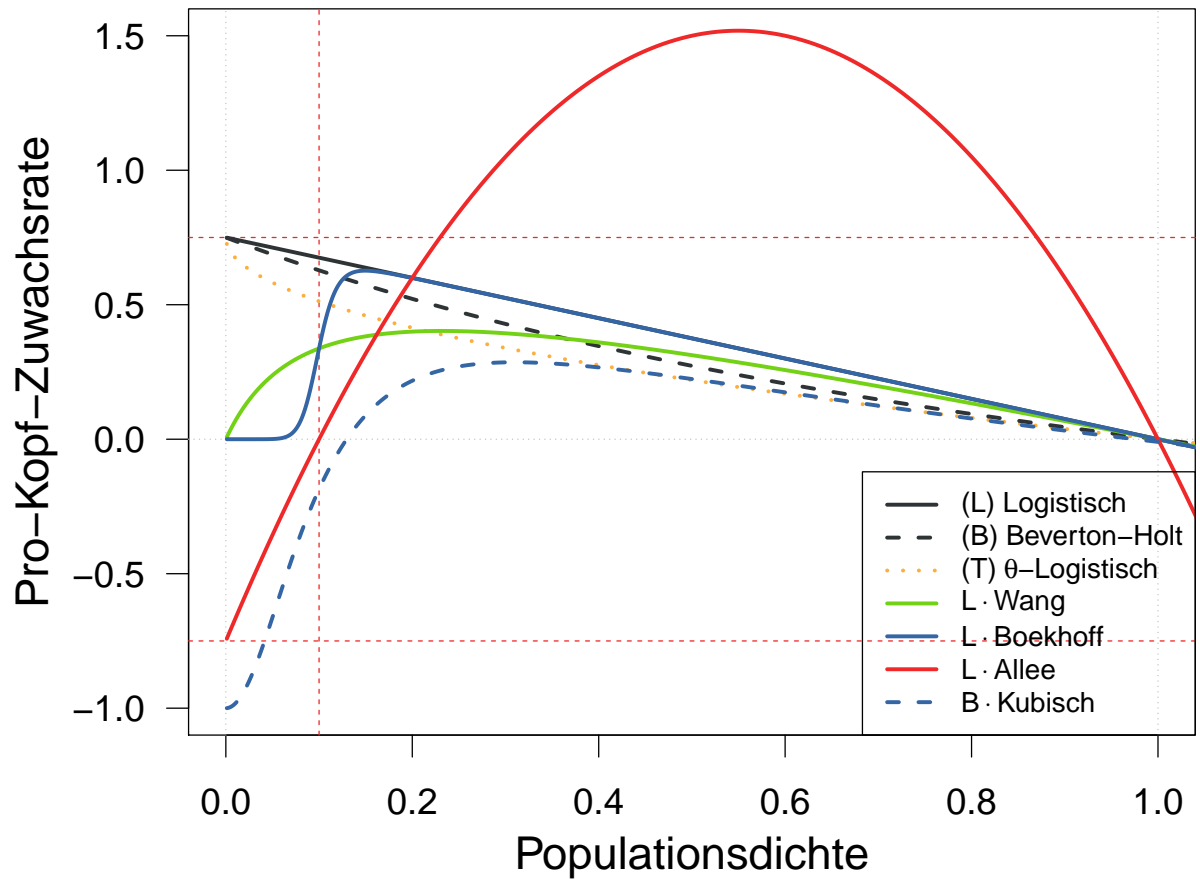


Abbildung A.5.: **Unterschiedliche dichteabhängige Wachstumsfunktionen.**

Dargestellt sind die Pro-Kopf-Zuwachsrates dreier Funktionen für negativ-dichteabhängiges Wachstum (siehe TAB. A.1) sowie von vier Funktionen, in denen der Allee-Effekt mit einbezogen wurde (siehe dazu TAB. A.2). $r = 0.75$; $\theta = 0.5$; $\frac{K_C}{K} = \alpha = 0.1$; $C = \alpha * K$.

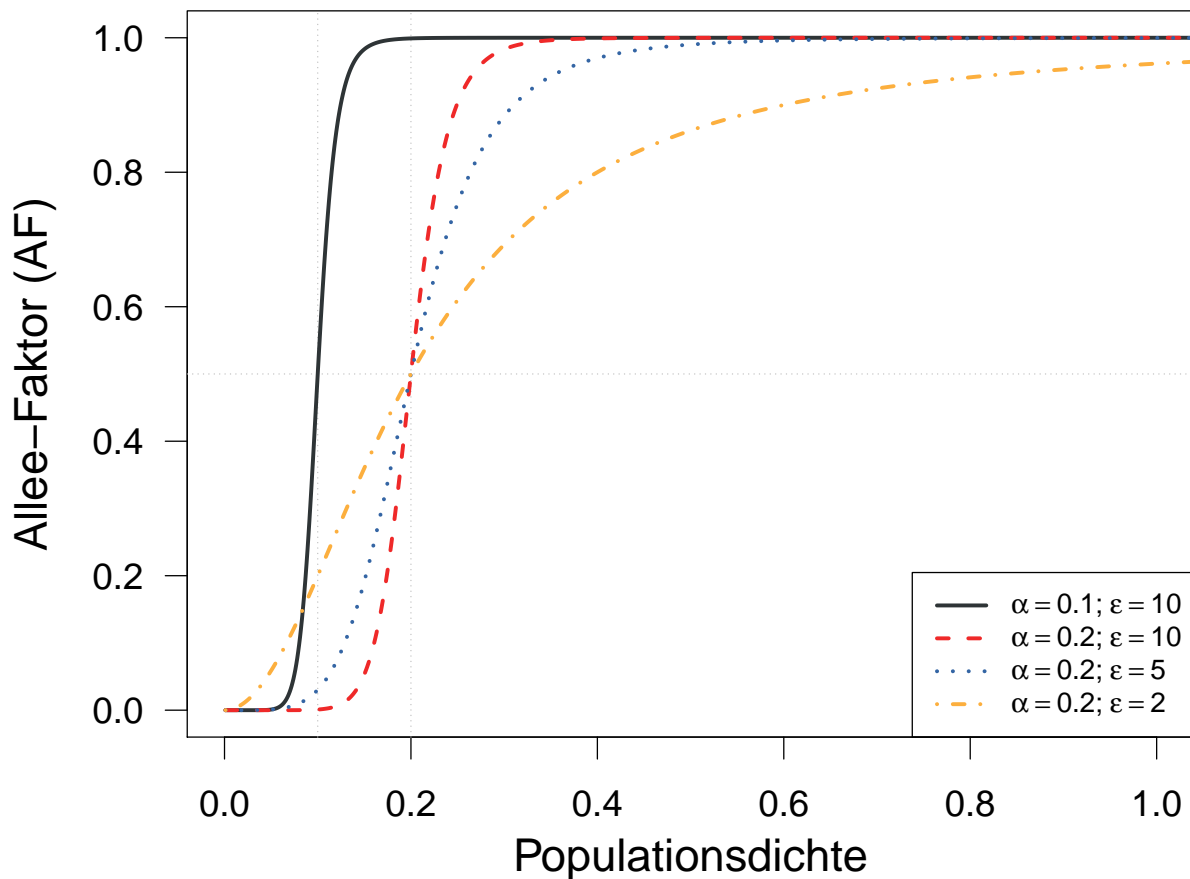


Abbildung A.6.: **Bedeutung der Parameter α und ϵ meiner Allee-Effekt Implementierung.** Der Wert für α bestimmt diejenige Populationsdichte, an der $AF = 0.5$ ist. Dadurch wird das Pro-Kopf-Wachstum an dieser Stelle halb-maximal. Der Wert für ϵ bestimmt die Breite des Bereiches, in dem AF zwischen „nahe Null“ und „nahe Eins“ liegt. Für $\alpha = n.a.$ oder $\epsilon = n.a.$ gab es keinen Allee-Effekt ($AF = 1$).

A. Anhang (Methoden)

Listing A.1: **Implementierung der Multinomialverteilung in Java.** Die Funktion `nextValue()` eines `MultinomialGenerator`-Objektes gibt einen Array zurück, der die tatsächliche Anzahl der Individuen jedes einzelnen Typs in der nächsten Generation enthält. Hierzu verwendet die Funktion den Vektor `probs`, der die erwarteten Wahrscheinlichkeiten der einzelnen Typen enthält. Solange die Stichprobengröße `sampleSize` noch nicht erreicht ist, wird dazu aus einer Binomialverteilung gezogen (Devroye, 1986). Die von mir verwendete Binomialverteilungs-Klasse stammt aus der Bibliothek *Uncommons Math* (Dyer, 2010).

```
package info.boekhoff.sven.random;

import java.util.Random;
import org.uncommons.maths.random.BinomialGenerator;

/** Wrapper für das Ziehen aus einer Multinomialverteilung. */
public class MultinomialGenerator {

    /** Zufallszahlen Generator (z.B. MersenneTwister). */
    private Random randomNumberGenerator;
    /** Vektor mit erwarteten Wahrscheinlichkeiten. */
    private double[] probs;
    /** Summe der Wahrscheinlichkeiten (in der Regel 1). */
    private double probSum;

    /** Konstruktor. */
    public MultinomialGenerator(final double[] initialFrequencies, final Random rng) {
        // Referenz auf den Zufallszahlengenerator:
        this.randomNumberGenerator = rng;
        // Referenz auf die erwarteten Wahrscheinlichkeiten:
        this.probs = initialFrequencies;
        // Summe der erwarteten Wahrscheinlichkeiten:
        for (double d : this.probs) {
            this.probSum += d;
        }
    } // ENDE: Konstruktor

    /** Stichprobe der Größe 'sampleSize' aus der Multinomialverteilung ziehen */
    public final int[] nextValue(final int sampleSize) {
        /** Ergebnis-Vektor. */
        int[] result = new int[this.probs.length];
        /** Arbeitskopie der Wahrscheinlichkeitsumme. */
        double total = this.probSum;
        /** Arbeitskopie der Stichprobengröße. */
        int remainingSize = sampleSize;
        /** Referenz auf aktuelle Binomialverteilung. */
        BinomialGenerator gen;

        // Für jede Typen-Wahrscheinlichkeit
        for (int i = 0; i < this.probs.length; i++) {
            // Sofern die Stichprobengröße noch nicht erreicht ist
            if (remainingSize > 0) {
```



```

    if ((this.probs[i] / total) >= 1) {
        // Die Stichprobe besteht vollständig aus dem aktuellen Typen,
        // falls die Wahrscheinlichkeit aller verbleibenden Typen 0 ist
        result[i] = remainingSize;
        remainingSize = 0;
        total = 0;
    } else if (this.probs[i] > 0) {
        // Erzeuge eine Binomialverteilung
        gen = new BinomialGenerator(remainingSize, (this.probs[i] / total),
            this.randomNumberGenerator);
        // Ziehe eine Stichprobe aus der Verteilung
        int sample = gen.nextValue();
        result[i] = sample;
        // Verringere die Zahl der noch zu ziehenden Individuen
        remainingSize -= sample;
        // Verringere die Bezugswahrscheinlichkeit
        total -= this.probs[i];
    } else {
        // Wahrscheinlichkeit des betrachteten Typs ist 0
        result[i] = 0;
    }
} else {
    // Stichprobengröße ist erreicht
    result[i] = 0;
}
} // ENDE: for each type
return result;

} // ENDE: nextValue()

} // ENDE: MultinomialGenerator

```

Listing A.2: Implementierung der multi-hypergeometrischen Auswahl in Java.

Um den Speicherverbrauch gering zu halten, ist nicht jedes Individuum gesondert im Speicher vertreten. Stattdessen enthält jedes Dem einen Array mit den Individuenzahlen jedes Typs (*this.individuen[]*). Die Funktion *selectNewHypergeometricEmigrant()* wählt aus dem Array *this.individuen[]* ein Individuum für die Migration aus und gibt dessen Typ zurück. Dabei ändert sich sowohl die Zusammensetzung der Population, als auch die Zahl der Individuen, die im Dem verbleiben (*Ziehen ohne Zurücklegen*).

```

package info.boekhoff.sven.expander.demes;
/**
 * Zieht ein einzelnes Individuum für die Migration.
 * Dabei ändert sich die Zusammensetzung des Demes.
 */
private short selectNewHypergeometricEmigrant(final RandomSven zufall) {
    // Zufallszahl: Welches Individuum soll migrieren?
    short zufallszahl = (short)
        (zufall.getRandomSequence().nextInt(this.zahlDerIndividuenImDeme) + 1);
}

```

A. Anhang (Methoden)

```
// Von welchem Typ (keyEnd) ist dieses Individuum?  
// = In welchem Teil des Typen-Arrays befindet es sich?  
short summation = 0;  
short keyEnd = -1;  
while (summation < zufallszahl) {  
    keyEnd++;  
    summation += this.individuen[keyEnd];  
}  
  
// Ändere die Zusammensetzung des Dems!  
this.individuen[keyEnd]--;  
this.zahlDerIndividuenImDeme--;  
// Gib den Typen des migrierenden Individuums zurück  
return keyEnd;  
}
```

A.3. Auswertungen

A.3.1. Allelfrequenz-Spektren

A.3.1.1. Einführung

Die Simulationen von Range-Expansions und anderen demographischen Szenarien habe ich in der Regel 10^6 mal durchgeführt. Alle 10 Generationen habe ich die Frequenz p_f des abgeleiteten Allels gemessen. In Allelfrequenz-Spektren stelle ich den Anteil der Simulationen dar, bei denen zum Zeitpunkt g der Simulation das abgeleitete Allel eine gegebene Frequenz p_f im neu besiedelten Bereich hat – z.B. 'Nach 400 Generationen hat das abgeleitete Allel in 10% der Simulationen eine Frequenz von $p_f = 0.25\%$ '. Die Spektren stellen somit die Wahrscheinlichkeit dar, dass das abgeleitete Allel zum Zeitpunkt g einer Simulation die Frequenz p_f hat.

Individuelle Frequenzen p_f kommen selten vor. Daher habe ich mehrere Frequenzen zu einer Klasse mit der Klassenmitte P_n zusammengefügt. Da nur in wenigen Simulationen die abgeleiteten Allele eine hohe Frequenz p_f erreichen, habe ich eine logarithmische Klassifizierung verwendet (Newman, 2005). Die Intervallränder für die Klassifizierung sind B^{-nF} (für $n = 0, 1, 2, \dots$; Basis $B > 1$; gewählter Exponent F ; siehe auch TEIL-ABSCHN. A.3.1.3). Die Implementierung der Klassengrenzen in R wird in CODE A.3 beschrieben. Jede Klasse identifiziere ich mit der zugehörigen Klassenmitte P_n :

$$P_n = (B^{-nF} + B^{-(n+1)F})/2.$$

Die Breite ΔP_n einer Klasse ist gegeben durch:

$$\Delta P_n = B^{-nF} - B^{-(n+1)F}.$$

Die Wahrscheinlichkeit dafür, dass eine Frequenz p_f in die Klasse P_n fällt, ist damit:

$$\hat{f}(P_n) = \mathbb{P}(p_f \in [B^{-nF}, B^{-(n+1)F}]).$$

Auf Grundlage dieser Notationen definiere ich nun die Häufigkeitsdichte:

$$f(P_n) = \frac{\hat{f}(P_n)}{\Delta P_n}. \quad (\text{A.10})$$

Da die Klassenmitten P_n Klassen von gemessenen Allelfrequenzen repräsentieren, schreibe ich im Hauptteil der Arbeit der Einfachheit halber $f(p_f)$ anstatt von $f(P_n)$.

A.3.1.2. Fehlerrechnung

Zu jedem meiner Szenarien existiert eine Wahrscheinlichkeitsverteilung von Allelfrequenzen. Durch die Simulationen ermittle ich eine Stichprobe dieser Verteilungen. Um den durch die Unvollständigkeit entstehenden Fehler abschätzen zu können, verwende ich eine Resampling-Methode, das Bootstrapping.

Die Annahme beim Bootstrapping ist, dass man bereits die Grundgesamtheit abgebildet hat. Die Frage ist nun, wie genau ein Wert ist, wenn wir wiederholt aus dieser Grundgesamtheit ziehen würden. Dazu kann man die Schwankung des Mittelwertes (Standardfehler) der $\hat{f}(P_n)$ -Werte über die Bootstrap-Zyklen berechnen.

Zunächst benötigen wir dazu einen Schätzer Y^n (*estimator*) für $\hat{f}(P_n)$ der Klasse n :

$$Y^n \equiv \frac{1}{N} \cdot \sum_{m=1}^N X_m^n. \quad (\text{A.11})$$

Y^n schätzt die Wahrscheinlichkeit, bei N Simulationen eine Frequenz in der Klasse n zu erhalten. X_m^n ist ein Schätzer für die m -te Realisation eines Zufallsprozesses, bei dem geschaut wird, ob die Frequenz in der Klasse n liegt ($X_m^n = 1$) oder nicht ($X_m^n = 0$). Der Erwartungswert von X_m^n entspricht dabei dem gemessenen Wert unserer Verteilung für die Klasse n , $\hat{f}(P_n)$:

A. Anhang (Methoden)

$$\mathbb{E}[X_m^n] = \hat{f}(P_n). \quad (\text{A.12})$$

Auch der Erwartungswert für den Schätzer von $\hat{f}(P_n)$, Y^n , entspricht dem gemessenen Wert für die entsprechende n -te Klasse, also $\hat{f}(P_n)$:

$$\begin{aligned} \mathbb{E}[Y^n] &= \frac{1}{N} \cdot \sum_{m=1}^N \mathbb{E}[X_m^n] \quad | \text{ mit GL. A.12} \\ &= \frac{1}{N} \cdot \sum_{m=1}^N \hat{f}(P_n) \\ &= \frac{1}{N} \cdot N \cdot \hat{f}(P_n) \\ &= \hat{f}(P_n). \end{aligned}$$

Zur Berechnung des Standardfehlers der Häufigkeitsdichte benötigen wir die Varianz $\text{Var}[f(P_n)]$ (siehe auch GL. A.10):

$$\begin{aligned} \text{Var}[f(P_n)] &= \text{Var}\left[\frac{Y^n}{\Delta P_n}\right] \quad | \text{ mit GL. A.14} \\ &= \frac{1}{\Delta P_n^2} \cdot \text{Var}[Y^n]. \end{aligned} \quad (\text{A.13})$$

Hinweis: Beim Rechnen mit Varianzen werden konstante Vorfaktoren im Quadrat vorangestellt:

$$\text{Var}[aX] = a^2 \cdot \text{Var}[X]. \quad (\text{A.14})$$

Hinweis: Die Summe der Varianzen zweier Zufallsvariablen entspricht der Varianz ihrer Summe. Sofern die Variablen unabhängig sind, ist die Kovarianz Null:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]. \quad (\text{A.15})$$

Nun wird die Varianz von Y^n benötigt (siehe auch GL. A.11):

$$\begin{aligned} \text{Var}[Y^n] &= \text{Var}\left[\frac{1}{N} \cdot \sum_{m=1}^N X_m^n\right] \quad | \text{ mit GL. A.14} \\ &= \frac{1}{N^2} \cdot \text{Var}\left[\sum_{m=1}^N X_m^n\right] \quad | \text{ Summe auflösen} \end{aligned} \quad (\text{A.16})$$

$$\begin{aligned} &= \frac{1}{N^2} \cdot N \cdot \text{Var}[X_m^n] \\ &= \frac{1}{N} \cdot \text{Var}[X_m^n]. \end{aligned} \quad (\text{A.17})$$

Da die Varianz innerhalb der Klasse n für alle einzelnen Realisierungen gleich ist (eine Konstante), kann das Summenzeichen in GL. A.16 aufgelöst werden (Hinweis: $\sum_{i=0}^N K = N \cdot K$; mit $K = \text{konstant}$).

Da es sich beim Bootstrapping um eine Methode „Ziehen mit Zurücklegen“ handelt, entspricht das Ergebnis einer binomialverteilten Zufallsvariablen. Die Varianz eines solchen Zufallsprozesses lässt sich folgendermaßen berechnen:

$$\begin{aligned} \text{Var}[X_m^n] &= \hat{f}(P_n) \cdot (1 - \hat{f}(P_n)) \cdot \underbrace{N}_{=1} \\ &= \hat{f}(P_n) \cdot (1 - \hat{f}(P_n)). \end{aligned} \quad (\text{A.18})$$

Die Wahrscheinlichkeit, die Frequenz des abgeleiteten Allels p_f nach der Simulation in der n -ten Klasse zu finden, ist gegeben durch $\hat{f}(P_n)$. Die Wahrscheinlichkeit, in einer anderen Klasse zu sein, ist $(1 - \hat{f}(P_n))$. Bei X_m^n handelt es sich um einen Schätzer für eine einzelne Simulation (siehe Erläuterungen zu GL. A.11). Daher ist die Zahl der Versuche $N = 1$ Simulation. Ziel ist aber, die Varianz über viele Simulationen zu ermitteln. Dazu wird GL. A.18 in GL. A.17 eingesetzt:

$$\begin{aligned} \text{Var}[Y^n] &= \frac{1}{N} \cdot \hat{f}(P_n) \cdot \underbrace{(1 - \hat{f}(P_n))}_{\approx 1} \\ &\approx \frac{1}{N} \cdot \hat{f}(P_n). \end{aligned} \quad (\text{A.19})$$

Da es selten vorkommt, dass ein abgeleitetes Allel sehr häufig wird, gilt für die Klassen

A. Anhang (Methoden)

hoher Allelfrequenzen $\hat{f}(P_n) \approx 0$. Daher gilt auch $1 - \hat{f}(P_n) \approx 1$.

Die Varianz der Wahrscheinlichkeitsdichte einer n -ten Klasse kann nun durch Einsetzen von GL. A.19 in GL. A.13 berechnet werden.

$$\begin{aligned}\text{Var}[f(P_n)] &= \frac{1}{\Delta P_n^2} \cdot \frac{1}{N} \cdot \hat{f}(P_n) && | \text{ mit GL. A.10} \\ &= \frac{1}{\Delta P_n^2} \cdot \frac{1}{N} \cdot f(P_n) \cdot \Delta P_n \\ &= \frac{f(P_n)}{N \cdot \Delta P_n}.\end{aligned}\tag{A.20}$$

Der Standardfehler von $f(P_n)$ errechnet sich hier durch:

$$\begin{aligned}SE[f(P_n)] &= \sqrt{\text{Var}[f(P_n)]} && | \text{ mit GL. A.20} \\ &= \sqrt{\frac{f(P_n)}{N \cdot \Delta P_n}}.\end{aligned}$$

A.3.1.3. Logarithmische Klassifizierung

Bei der logarithmischen Klassifizierung (*logarithmic binning*) befindet sich die breiteste Klasse (*bin*) auf der rechten Seite (Newman, 2005). Die Breite der Klassen nimmt um den Faktor B^F von rechts nach links ab (B = Basis; F = gewählter Exponent). Hat die oberste Klassengrenze den Wert B^u (u = Exponent der obersten Klassengrenze, in der Regel $u = 0$), so ist die obere Klassengrenze der n -ten darunterliegenden Klasse bei:

$$B^u / B^{n \cdot F} = B^{u - F \cdot n}.\tag{A.21}$$

Meine Implementierung der Berechnung der Klassengrenzen (in R) ist in CODE A.3 dargestellt.

Listing A.3: Die Funktion `logbins()` berechnet Klassengrenzen (*bin breaks*), die für die logarithmische Klassifizierung (z.B. mit Hilfe der R-Funktion `hist()`) genutzt werden können. Die Errechnung der Klassengrößen kann auf zwei Weisen erfolgen: (A) Über die Angabe der Breite der größten Klasse (*widthOfLargestBin*). (B) Über den Exponenten F (*exponentOfDeclineFactor*) des Faktors, mit dem die Klassen kleiner werden (siehe GL. A.21).

```
logbins <- function(
  smallestValue           # Kleinste mögliche Frequenz
  , largestValue = 1     # Größte mögliche Frequenz (obere Grenze)
  , exponentOfDeclineFactor = 0.1 # Verkleinerungsfaktor
  , widthOfLargestBin = NULL # Breite der größten Klasse
  , base = exp(1)        # Basis des Log-Raumes (Standard: e)
) {

  # Exponent der größten Frequenz
  # (die oberste Klassengrenze ist: base ^ exponentOfUpperLimit)
  exponentOfUpperLimit = log(largestValue , base=base)

  # Exponent der kleinsten Frequenz
  exponentOfTheSmallestValue = log(smallestValue , base=base)

  # Errechnung des Verkleinerungsfaktors , wenn widthOfLargestBin gesetzt ist .
  if(!is.null(widthOfLargestBin)) {
    exponentOfDeclineFactor = -1 * log((1 - (widthOfLargestBin / largestValue)) ,
      base=base)
  }

  # Die obere Klassengrenze der n-ten Klasse ist :
  # base ^ exponentOfUpperLimit / base ^ (n * exponentOfDeclineFactor)
  # = base ^ (exponentOfUpperLimit - n * exponentOfDeclineFactor)
  #
  # Zunächst werden die oberen Grenzen aller Klassen bestimmt .
  # Zusätzlich wird die untere Grenzen der kleinsten Klasse bestimmt
  # (daher -exponentOfDeclineFactor) .
  #
  # Da der der kleinste Wert oft stark schwankt ,
  # würde die obere Grenze nicht immer am gleichen Punkt liegen .
  # Daher wird von der größten zur kleinsten Klasse gezählt ,
  # damit die obere Grenze festgehalten werden kann (Standard: exp(0) = 1) .
  #
  exponentRange = (exponentOfUpperLimit - exponentOfTheSmallestValue) +
    exponentOfDeclineFactor

  nTimesExponentOfDeclineFactor = seq(
    from = 0           # Vom größten Exponenten
    ,to = exponentRange # Einen weiteren Schritt nach links!
    ,by = exponentOfDeclineFactor # Die Exponenten werden immer kleiner => -
  )

  exponents = (exponentOfUpperLimit - nTimesExponentOfDeclineFactor)
```

A. Anhang (Methoden)

```
binBreaks = c(
  0, # Die Null sammelt alle verlorenen Allele auf!
  rev( base ^ exponents ) # Ich möchte von der Klasse der kleinsten Werte
  ) # ENDE: c()
  # zur Klasse der größten Werte gehen => rev()

# binBreaks zurückgeben
binBreaks

} # ENDE: logBins()
```

A.3.1.4. Zipf-Plots

In den Allelfrequenz-Spektren aus den zuvor beschriebenen klassifizierten Wahrscheinlichkeitsdichte-Verteilungen, war ein Potenzgesetz zu erkennen mit folgender Form:

$$f(p_f) = C \cdot p_f^{a_k}.$$

In Zipf-Plots wird die Wahrscheinlichkeit dargestellt, eine Frequenz größer als p_f zu erreichen (Newman, 2005):

$$F(p_f) = f(> p_f) = \frac{C}{a_k + 1} \cdot p_f^{a_k+1} = \frac{C}{a_z} \cdot p_f^{a_z}.$$

Für den in Zipf-Plots bestimmten Exponenten a_z gilt daher:

$$a_z = a_k + 1.$$

A.3.2. Echte genetische Daten

Für die Analyse echter genetischer Daten habe ich Ergebnisse des 1000-Genome-Projektes verwendet. Ziel des 1000-Genome-Projektes ist es, die vollständige Erbinformation von mehr als 2500 menschlichen Individuen aus 26 verschiedenen Populationen zu sequenzieren und zu analysieren. In einer Pilotphase wurden zunächst 180 Genome (1000 Genomes Project Consortium, 2010), und in der „Phase 1“ 1092 Genome sequenziert (1000 Genomes Project Consortium et al., 2012).

Für die in dieser Arbeit gezeigten Analysen habe ich ausschließlich Daten der „Phase 1“ verwendet (Release: 20110521; Datei: *ALL.wgs.phase1_release_v3.20101123.snps_indels_sv.sites.vcf*). Diese enthielten Informationen aus 14 Populationen von vier Kontinenten (siehe ABB. A.7). Die Allelfrequenz-Spektren habe ich aus Frequenzen von abgeleiteten Einzelnukleotid-Polymorphismen (SNP; *Single Nucleotide Polymorphism*) erstellt. Um dabei vor allem neutrale Allele zu verwenden, habe ich nur SNPs aus nicht kodierenden Regionen in die Auswertung einbezogen. Darüber hinaus habe ich diejenigen SNPs ausgeschlossen, für die nicht in allen sequenzierten Individuen Daten vorhanden waren oder für die das ursprüngliche Allel nicht (eindeutig) bestimmt werden konnte. Zudem habe ich Genorte ausgeschlossen, an denen mehr als ein alternatives Allel in der Population vorhanden war. Aus den Allelfrequenzen für das Referenz-Allel und der Information, welches Allel ursprünglich ist, habe ich die Frequenzen p_f des abgeleiteten Allels ermittelt.

Allelfrequenz-Spektren für **Standing-genetic-Variation** habe ich aus solchen Allelen berechnet, die sowohl in Afrika als auch (mindestens) in der betrachteten Population (Europa, Amerika bzw. Asien) zu finden waren.

Allelfrequenz-Spektren für **Ongoing-Mutations** habe ich aus solchen Allelen berechnet, die nur in einer einzigen kontinentalen Population zu finden waren. Diese Mutationen sind wahrscheinlich nach Beginn der Range-Expansion in den entsprechenden Populationen aufgetreten.

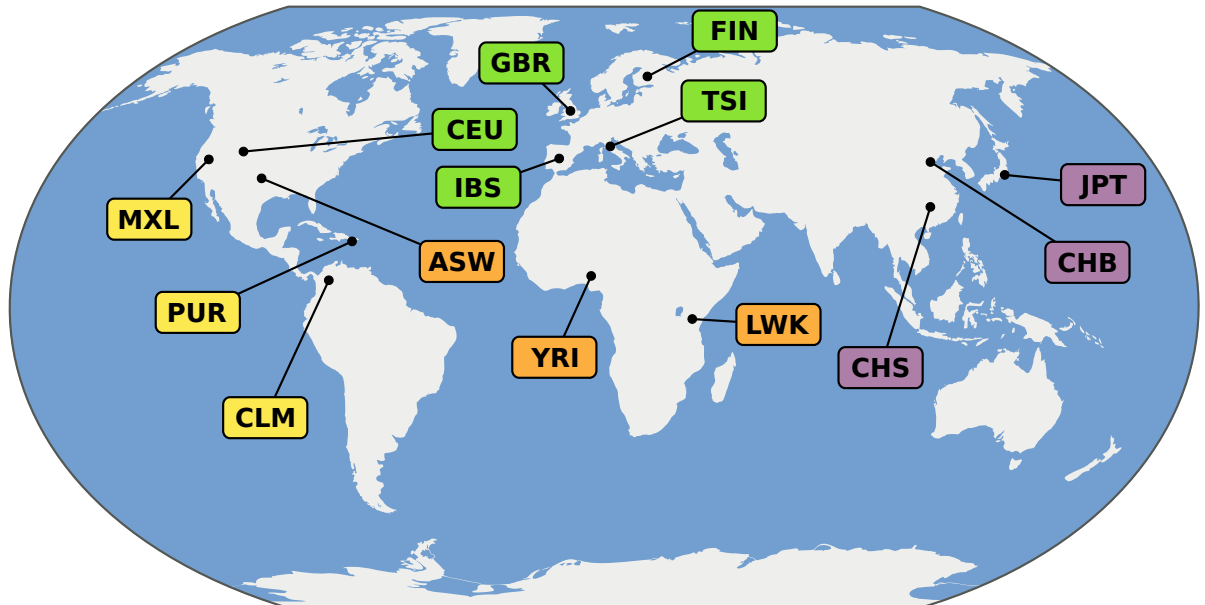


Abbildung A.7.: **Untersuchte Populationen in der „Phase 1“ des 1000-Genom-Projektes.** Die einzelnen Populationen sind zu vier kontinentalen Gruppen (Europa, Afrika, Asien und Amerika) zusammengefasst. Nachfolgend gebe ich für jede Population eine Beschreibung sowie die Zahl n der sequenzierten Individuen an. **AMR** (Amerika; $n = 181$): *MXL* (Bevölkerung mit mexikanischen Vorfahren in Los Angeles, Kalifornien, USA; $n = 66$), *CLM* (Kolumbianer in Medellín, Kolumbien; $n = 60$), *PUR* (Puerto-Ricaner in Puerto Rico; $n = 55$). **AFR** (Afrika; $n = 246$): *ASW* (Bevölkerung mit afrikanischen Vorfahren im Südwesten der USA; $n = 61$), *LWK* (Luhya aus Webuye, Kenia; $n = 97$), *YRI* (Yoruba aus Ibadan, Nigeria; $n = 88$); **ASN** (Asien; $n = 286$): *CHB* (Han Chinesen aus Peking, China; $n = 97$), *CHS* (Han Chinesen aus dem Süden Chinas; $n = 100$), *JPT* (Japaner aus Tokio, Japan; $n = 89$); **EUR** (Europa; $n = 379$): *CEU* (Einwohner Utahs, mit Vorfahren aus dem Norden und Westen Europas; $n = 85$), *FIN* (Finnen aus Finnland; $n = 93$), *GBR* (Briten aus England und Schottland; $n = 89$), *TSI* (Toskaner aus Italien; $n = 98$), *IBS* (Iberische Populationen aus Spanien; $n = 14$); Quelle: ([1000 Genomes Project Consortium et al., 2012](#), Abbildung S2 und Abschnitt 5.10). Bildvorlage: KAP. C.

B. Anhang (Ergebnisse)

Dieses Kapitel enthält zusätzliche Simulationsergebnisse (ABSCHN. B.1), Argumente für den theoretisch erwarteten Exponenten des Potenzgesetzes (ABSCHN. B.2) sowie Ergebnisse aus der Analyse echter genetischer Daten (ABSCHN. B.3).

B.1. Simulationsergebnisse

B.1.1. Vergleich demographischer Szenarien

Für das demographische Szenario „räumlich stabiles Wachstum“ habe ich die Wachstumsrate der Gesamtpopulation r_{deme} benötigt. In ABB. B.1 sind zwei Arten dargestellt, wie ich diese Wachstumsrate bestimmt habe. Dabei ist zu erkennen, dass das Wachstum der Population nahezu linear ist.

In ABB. B.2 und ABB. B.3 sind die Allelfrequenz-Spektren von Standing-genetic-Variation bzw. Ongoing-Mutations für drei demographische Szenarien nach $g = 400$ Generationen dargestellt. Die Spektren der gleichen Simulationen nach $g = 1990$ Generationen sind in ABB. 3.1 und ABB. 3.2 dargestellt.

B.1.2. Vergleich von Standing-genetic-Variation und Ongoing-Mutations

In ABB. B.4 zeige ich Allelfrequenz-Spektren für verschiedene Mutationsraten μ und verschiedene Anfangsfrequenzen p_s . Die normalisierten Spektren sind in ABB. 3.3 gezeigt.

In Grafik ABB. B.5 zeige ich die Abhängigkeit des Γ -Wertes von der Dauer der Range-Expansion. Dabei ist zu erkennen, dass die Bedeutung von Ongoing-Mutations für die Allelfrequenz-Spektren mit zunehmender Dauer der Range-Expansion steigt.

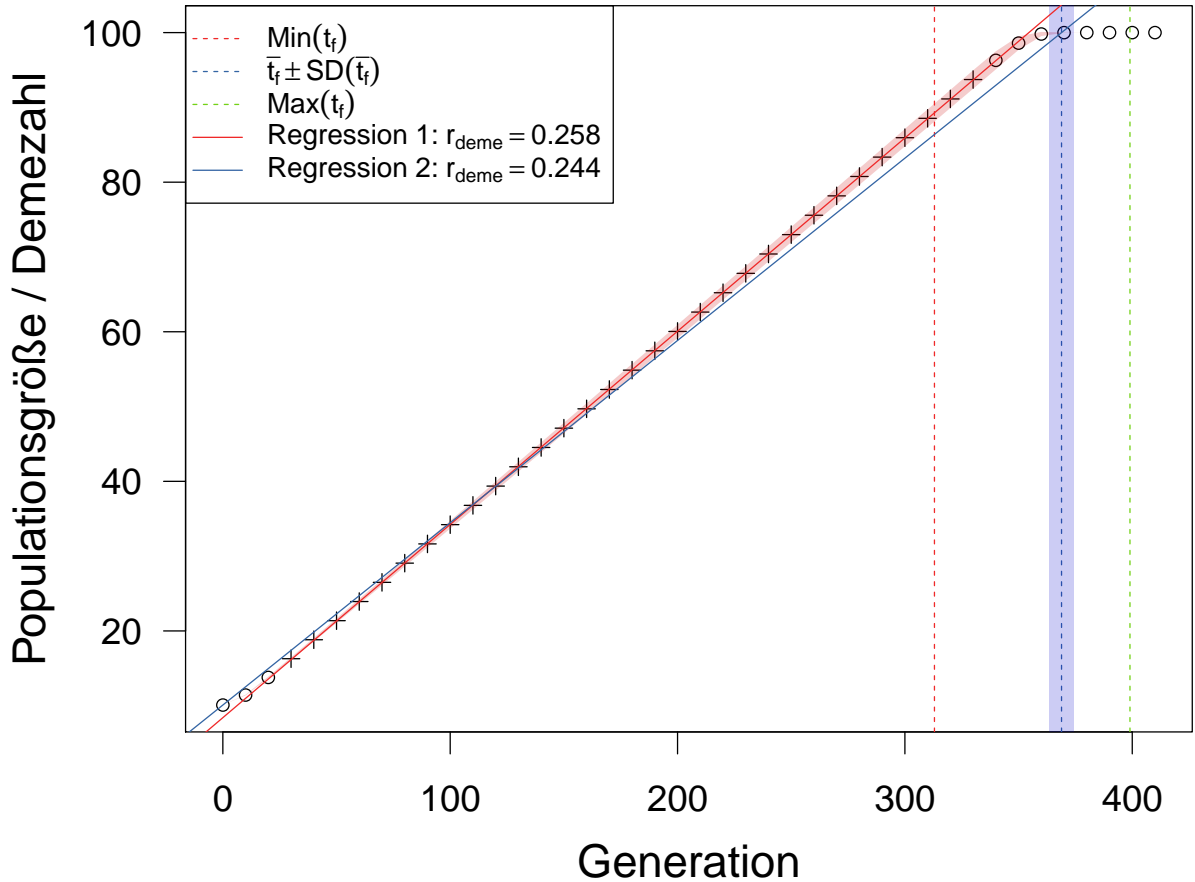


Abbildung B.1.: **Lineares Wachstum der Gesamtpopulation bei Range-Expansions.** Während das Wachstum innerhalb der Deme logistisch reguliert ist, nimmt die Anzahl der Individuen in der Gesamtpopulation näherungsweise linear zu. Gezeigt ist die mittlere Individuenzahl pro Dem ($K = 100$) aus 10^6 Simulationen in der Kunstwelt (rote Fläche: Standardabweichung). Die Zunahme der Kapazitätsgrenze für die Simulation des *räumlich stabilen Wachstums* ergibt sich aus der Steigung r_{deme} der Geraden. Die durchgezogene rote Gerade zeigt die lineare Interpolation derjenigen Werte, die durch Kreuze eingezeichnet sind. Die durchgezogene blaue Gerade liegt zwischen dem Startpunkt der Simulation ($t = 0$, $K_s = 10.19$) und dem Zeitpunkt \bar{t}_f (gestrichelte blaue Linie; blaue Fläche: Standardabweichung; siehe TAB. 3.2), zu dem im Mittel alle Deme vollständig besiedelt waren ($\bar{t}_f = 368.9$, $K = 100$). Zudem sind die kleinste und größte gemessene Laufzeit eingetragen (rote bzw. grüne gestrichelte Linie).

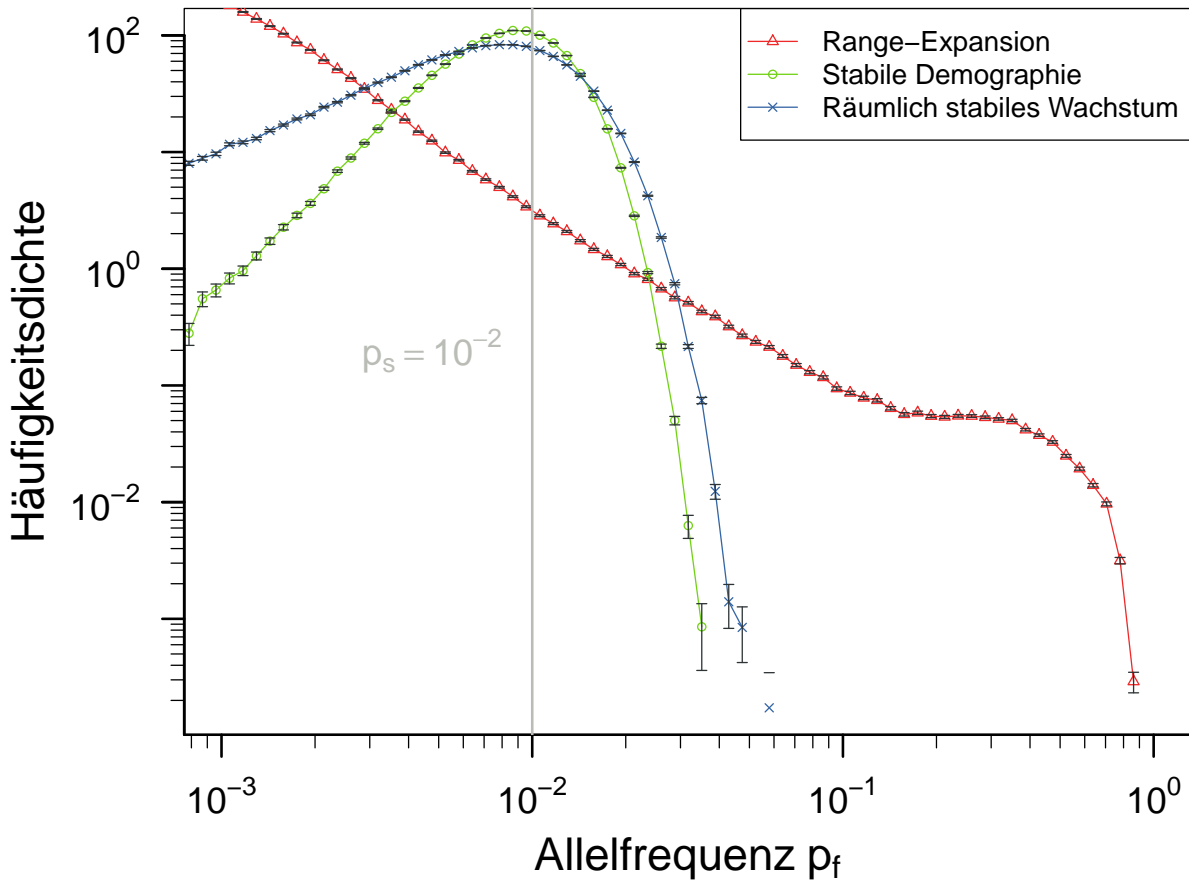


Abbildung B.2.: **Auswirkungen von drei demographischen Szenarien auf die Allelfrequenz-Spektren von Standing-genetic-Variation.** Die Allelfrequenz-Spektren wurden aus den gleichen Simulationsergebnissen erstellt, wie die Allelfrequenz-Spektren in ABB. 3.1, allerdings nach $g = 400$ Generationen. Das entspricht in etwa dem Zeitpunkt, zu dem die Range-Expansions abgeschlossen waren (siehe TAB. 3.2). Wie in ABB. 3.1 können anfänglich seltene genetische Varianten ($p_s = 10^{-2}$; vertikale graue Linie) während einer Range-Expansion (rote Dreiecke) deutlich höhere Frequenzen ($p_f^{max} = 0.88$) im neu besiedelten Bereich erreichen als in Simulationen einer stabilen Demographie (grüne Kreise; $p_f^{max} = 0.036$) oder räumlich stabilem Wachstum (blaue Kreuze; $p_f^{max} = 0.057$). Für weitere Informationen zu den Simulationen siehe ABB. 3.1.

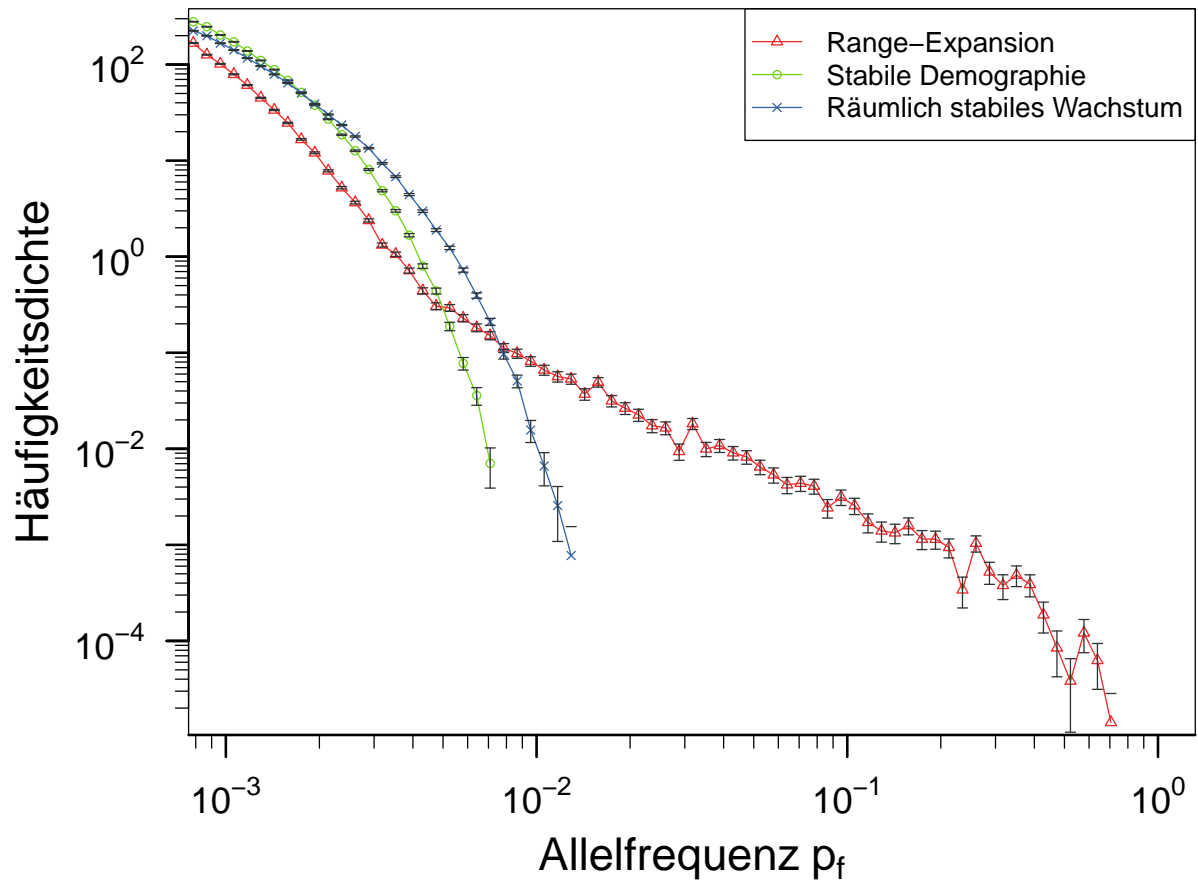


Abbildung B.3.: **Auswirkungen von drei demographischen Szenarien auf die Allelfrequenz-Spektren von Ongoing-Mutations.** Die Allelfrequenz-Spektren wurden aus den gleichen Simulationsergebnissen erstellt wie die Allelfrequenz-Spektren in ABB. 3.2, allerdings nach $g = 400$ Generationen. Das entspricht in etwa dem Zeitpunkt, zu dem die Range-Expansions abgeschlossen waren (siehe TAB. 3.2). Wie in ABB. 3.2 können neu auftretende Mutationen (Mutationsrate $\mu = 10^{-6}$) während einer Range-Expansion (rote Dreiecke) deutlich höhere Frequenzen ($p_f^{max} = 0.73$) im neu besiedelten Bereich erreichen als in Simulationen einer stabilen Demographie (grüne Kreise; $p_f^{max} = 0.007$) oder räumlich stabilem Wachstum (blaue Kreuze; $p_f^{max} = 0.013$). Für weitere Informationen zu den Simulationen siehe ABB. 3.2.

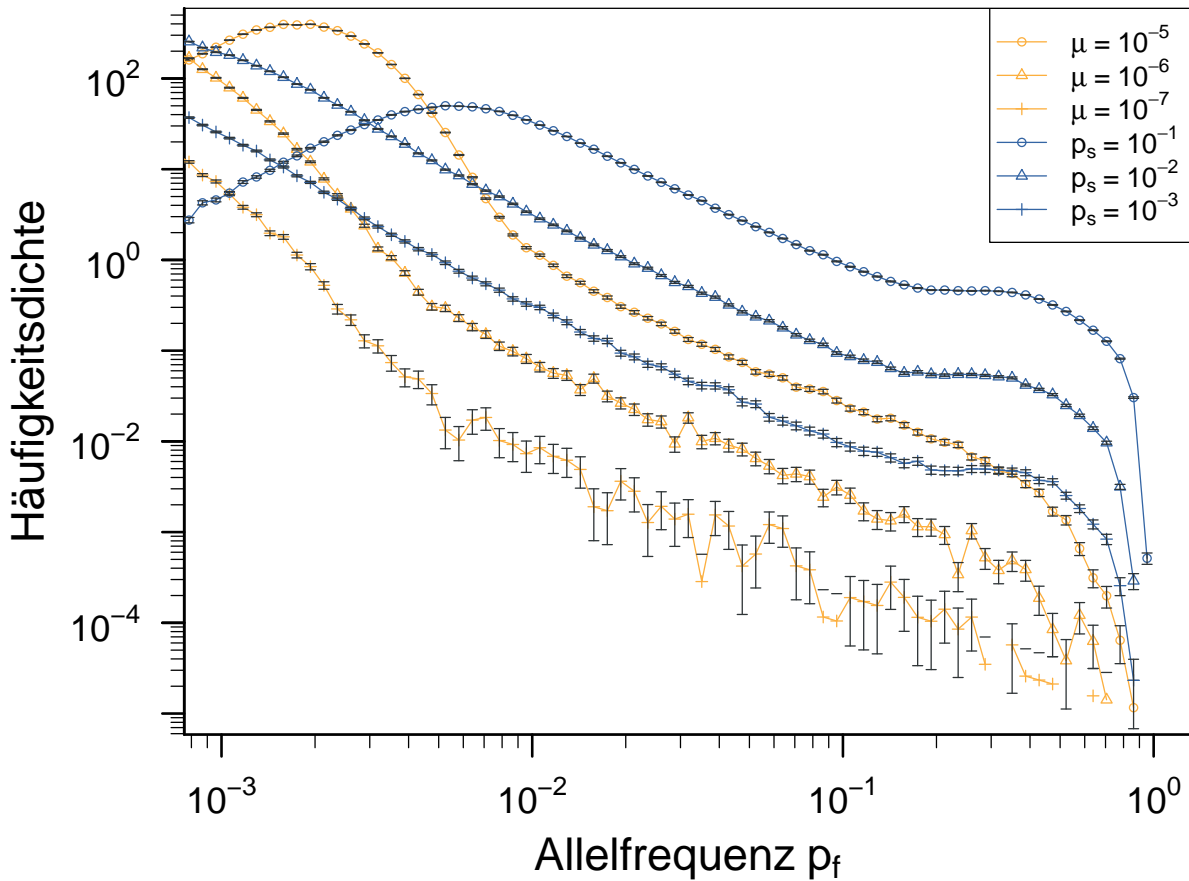


Abbildung B.4.: **Allelfrequenz-Spektren aus Range-Expansion Simulationen für Standing-genetic-Variation und Ongoing-Mutations.** Im Gegensatz zu ABB. 3.3 sind die Allelfrequenz-Spektren nicht zur Deckung gebracht worden. Jedes Spektrum wurde aus einer Million Allelfrequenzen von Simulationen auf der Kunstwelt erstellt. Die Messung erfolgte jeweils nach $g = 400$ Generationen, also direkt nach der Range-Expansion (siehe TAB. 3.2). Die Fehlerbalken geben den Standardfehler beim Bootstrapping der Verteilung an (siehe TEILABSCHN. A.3.1.2).

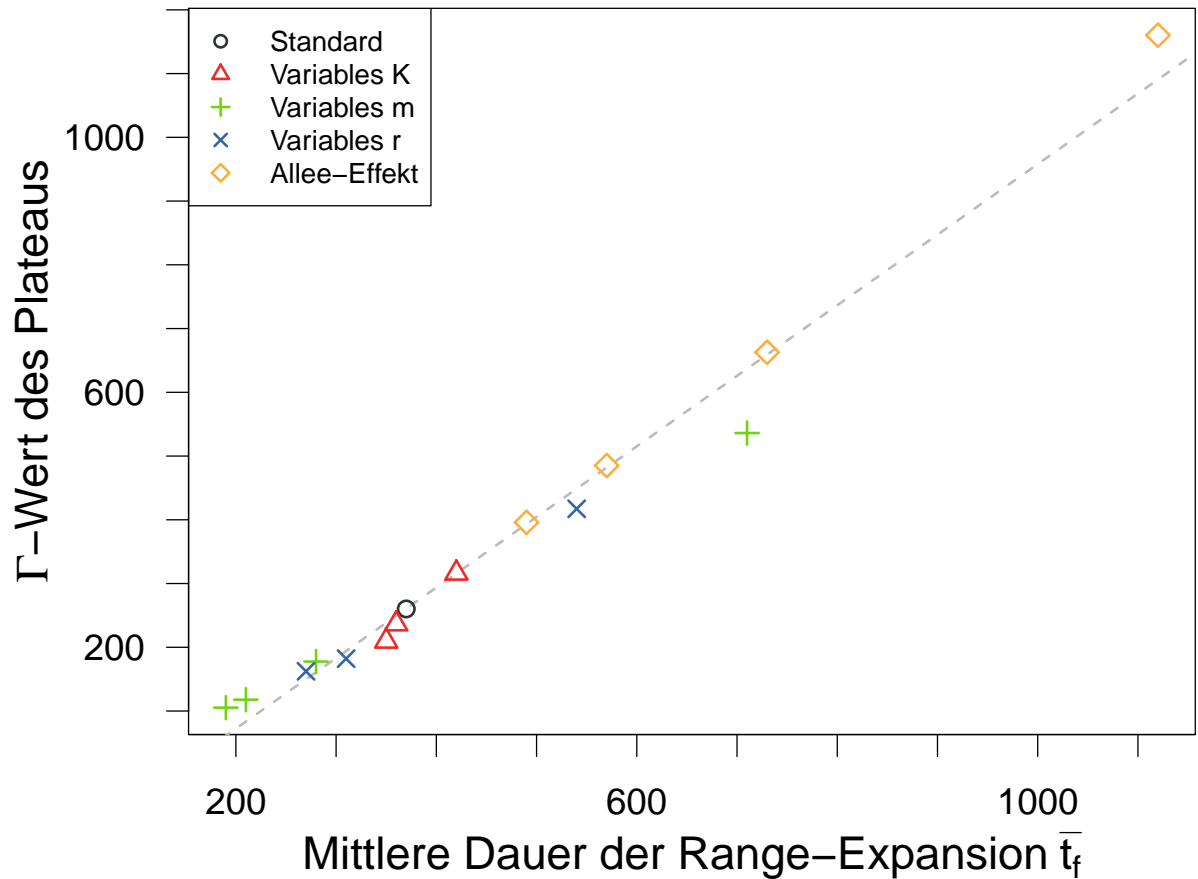


Abbildung B.5.: **Abhängigkeit des Γ -Wertes von der Dauer der Range-Expansion.** In erster Näherung ist der Wert für Γ direkt proportional zur mittleren Dauer \bar{t}_f (in Generationen) der Range-Expansion ($R^2 = 0.98$). Mit steigendem Γ -Wert nimmt auch die Bedeutung von Ongoing-Mutations für das Allelfrequenz-Spektrum zu (siehe TAB. 3.1). Dieser Anstieg lässt sich damit erklären, dass während einer längeren Range-Expansion mehr Mutationen an der Front auftreten und im Anschluss zu hohen Frequenzen surfen können. Die verwendeten Daten stammen aus TAB. 3.2.

Tabelle B.1.: Durchschnittliche Anzahl von Generationen, nach denen ein Individuum die letzte Reihe des Gitters erreicht hat (\bar{t}_e) bzw. alle Deme bis zur Kapazitätsgrenze besiedelt waren (\bar{t}_f). Die Werte wurden aus 20'000 Simulationen auf der Europakarte für verschieden starke Allee-Effekte ermittelt. In der Spalte „Parameter“ sind solche Werte angegeben, die sich von den Standardwerten ($m = 0.2$, $K = 100$, $r = 0.5$, $\alpha = n.a.$, $\epsilon = n.a.$; siehe auch TAB. 2.3) unterscheiden. (SD = Standardabweichung) Die Laufzeiten für die Kunstwelt sind in TAB. 3.2 aufgeführt.

Parameter	\bar{t}_e	$SD[t_e]$	\bar{t}_f	$SD[t_f]$
Standard Parameter	320.7	6.64	414.5	8.79
$\alpha = 0.05$, $\epsilon = 2$	427.1	5.63	534.3	8.05
$\alpha = 0.10$, $\epsilon = 10$	499.7	5.44	615.8	8.19
$\alpha = 0.20$, $\epsilon = 10$	645.4	7.05	792.9	10.94
$\alpha = 0.40$, $\epsilon = 10$	992.4	13.33	1247.9	21.11

B.1.3. Einfluss der Geographie auf die Allelfrequenz-Spektren

In TAB. B.1 sind die Laufzeiten für Simulationen auf der Europakarte eingetragen. Laufzeiten für Simulationen in der Kunstwelt finden sich in TAB. 3.2 auf Seite 43.

In ABB. B.6 sind die die normalisierten Allelfrequenz-Spektren für die Simulationen von Range-Expansions auf der Europakarte für verschiedene Mutationsraten μ und verschiedene Anfangsfrequenzen p_s dargestellt. Eine vergleichbare Grafik für Simulationen auf der Kunstwelt befindet sich auf Seite 39 (ABB. 3.3).

In ABB. B.7 sind die $\Gamma(p_f)$ -Spektren für die Simulationen von Range-Expansions auf der Europakarte zu zwei verschiedenen Zeitpunkten dargestellt (vergleiche ABB. 3.4 und ABB. 3.6).

B.1.4. Potenzgesetz in Simulationen auf der Großen Welt

In ABSCHN. 3.5 habe ich Allelfrequenz-Spektren aus Simulationen auf der Großen Welt für variable Migrationsraten m (ABB. 3.7) und verschieden starke Allee-Effekte (ABB. 3.8) gezeigt. In ABB. B.8 und ABB. B.9 zeige ich nun die Spektren für verschiedene Wachstumsraten r und unterschiedlich große Kapazitätsgrenzen K .

Den Exponenten des Potenzgesetzes habe ich durch lineare Regression in der doppelt-logarithmischen Darstellung der Allelfrequenz-Spektren ermittelt. Beispiele für die Bestimmung der Steigung in Dichteverteilungen klassifizierter Daten sowie kumulierter Allelfrequenz-Spektren (Zipf-Plots) sind in ABB. B.11 und ABB. B.12 gezeigt.

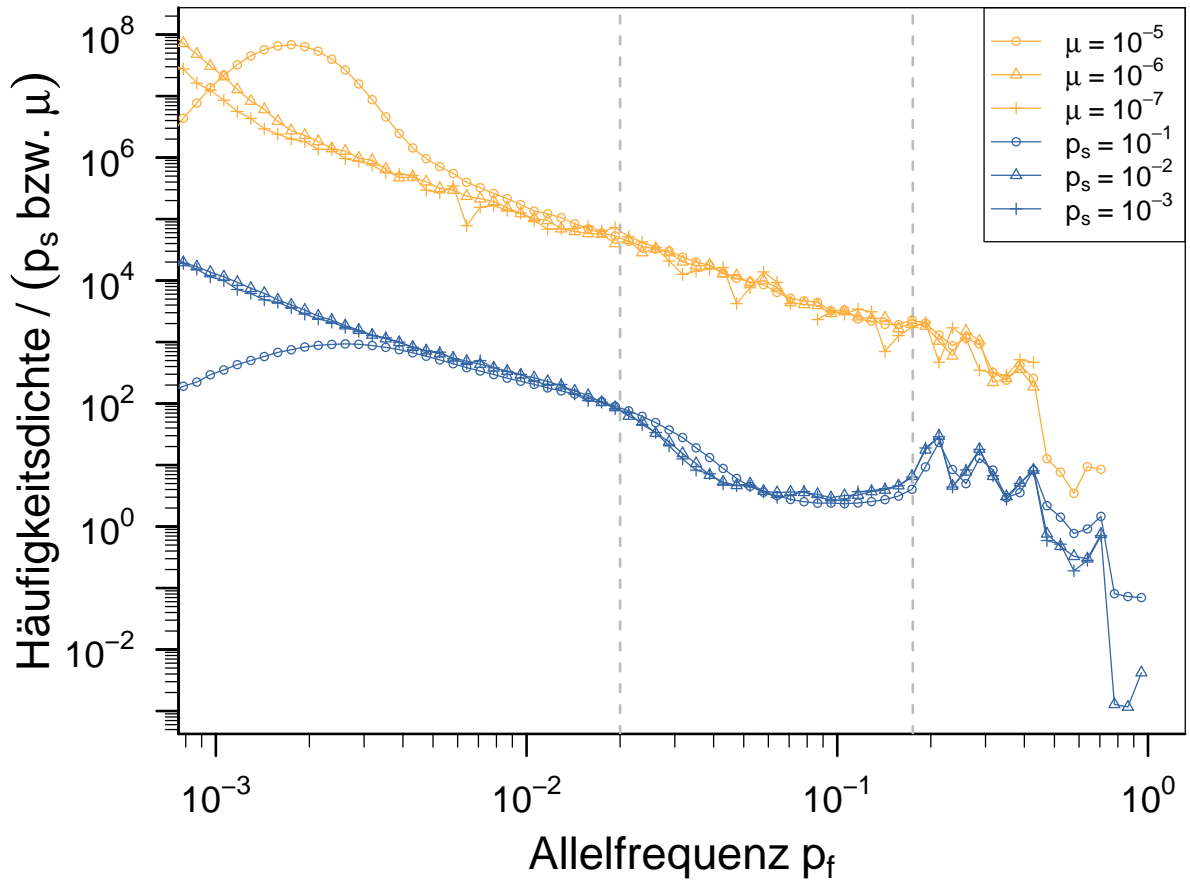


Abbildung B.6.: **Normalisierte Allelfrequenz-Spektren aus Range-Expansion Simulationen auf der Europakarte.** Die Allelfrequenz-Spektren können genauso wie die Spektren aus Simulationen auf der Kunstwelt (ABB. 3.3) zur Deckung gebracht werden, wenn man die Häufigkeitsdichten durch die verwendete Anfangsfrequenz p_s bzw. die Mutationsrate μ teilt (*Normalisierung*). Wie bereits in ABB. 3.5 beschrieben haben die geographischen Hindernisse vor allem Auswirkungen auf die Allelfrequenz-Spektren von Standing-genetic-Variation (blaue Kurve). Einige Frequenzen werden durch die Hindernisse unwahrscheinlicher (Bereich zwischen den gestrichelten Linien), da die Mutationen es entweder durch die Engstelle schaffen und in der Regel häufig werden (Bereich ganz rechts) oder es nicht schaffen und selten bleiben (Bereich ganz links). Die Allelfrequenzen für die gezeigten Spektren wurden aus jeweils 10^6 Simulationen nach $g = 400$ Generationen gemessen, also direkt nach der Range-Expansion (siehe TAB. 3.2).

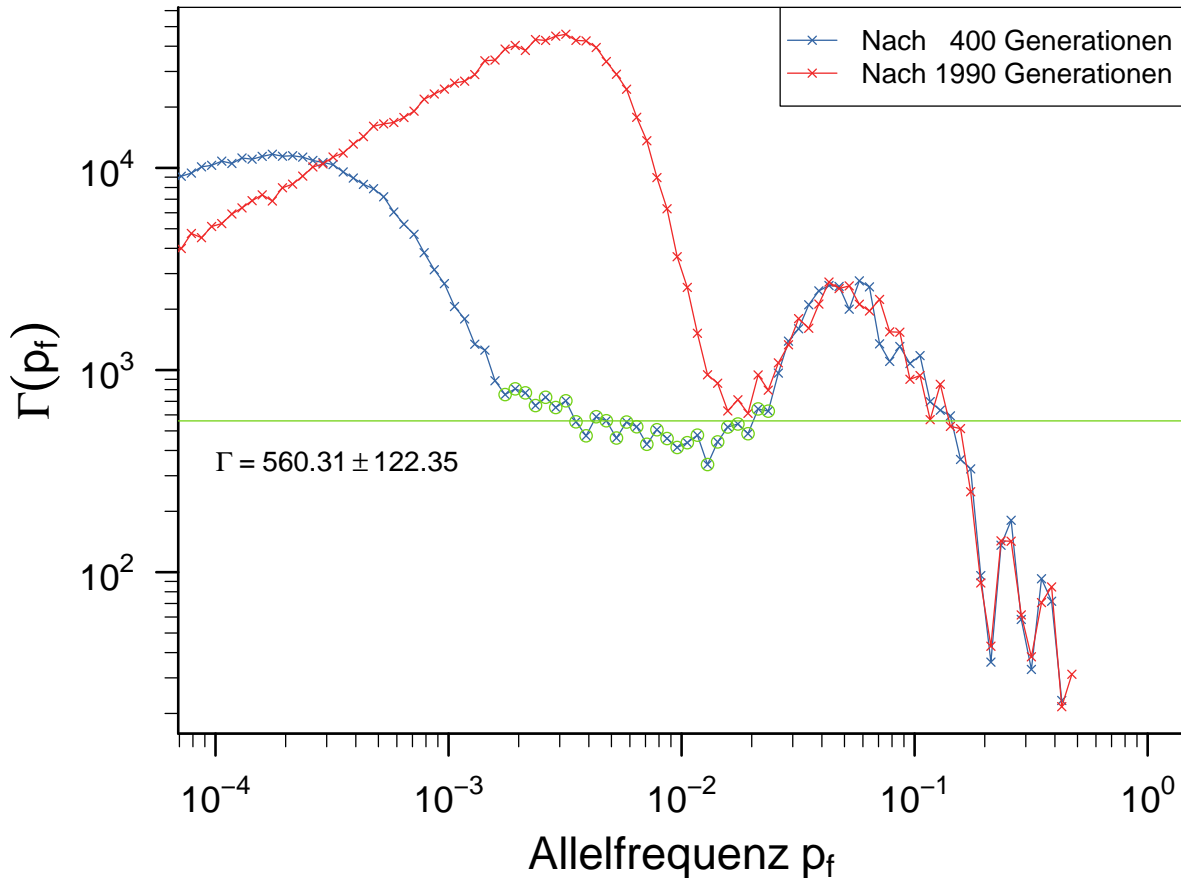


Abbildung B.7.: **Bedeutung von Ongoing-Mutations für das Allelfrequenz-Spektrum während und nach einer Range-Expansion auf der Europakarte.** Gezeigt ist der Quotient $\Gamma(p_f)$ aus den normalisierten Allelfrequenz-Spektren von Standing-genetic-Variation und Ongoing-Mutations. Ein steigender Wert für $\Gamma(p_f)$ deutet auf einen steigenden Einfluss von Ongoing-Mutations für die entsprechenden Allelfrequenzen p_f hin. Wie in ABB. 3.4 diskutiert spielen neu entstehende Mutationen für den Bereich niedriger Allelfrequenzen eine zunehmende Rolle. Darüber hinaus gibt es einen Bereich, in dem Allele aus Standing-genetic-Variation und Ongoing-Mutations gleichermaßen von der Range-Expansion betroffen sind (grün eingekreiste Punkte; diese wurden auch für die Berechnung des Mittelwertes verwendet). Verursacht durch die geographische Struktur gibt es (im Vergleich zu ABB. 3.4) Frequenzen im Bereich um $p_f = 5 \times 10^{-2}$, die vor allem aus Ongoing-Mutations stammen (Diskussion dazu siehe ABB. 3.6 und ABB. 3.5). Im Laufe der Zeit steigt der Anteil von Ongoing-Mutations im Bereich niedriger Allelfrequenzen, während es im Bereich hoher Frequenzen (hier: $p_f > 2 \times 10^{-2}$) kaum Änderungen über die Zeit gibt. $\Gamma(p_f)$ wurde aus Simulationen für $\mu = 10^{-6}$ und $p_s = 10^{-3}$ direkt im Anschluss an die Range-Expansion ($g = 400$ Generationen) sowie nach $g = 1990$ Generationen berechnet. ($K = 100$; $m = 0.2$; $r = 0.5$; Gitter: *Europakarte*)

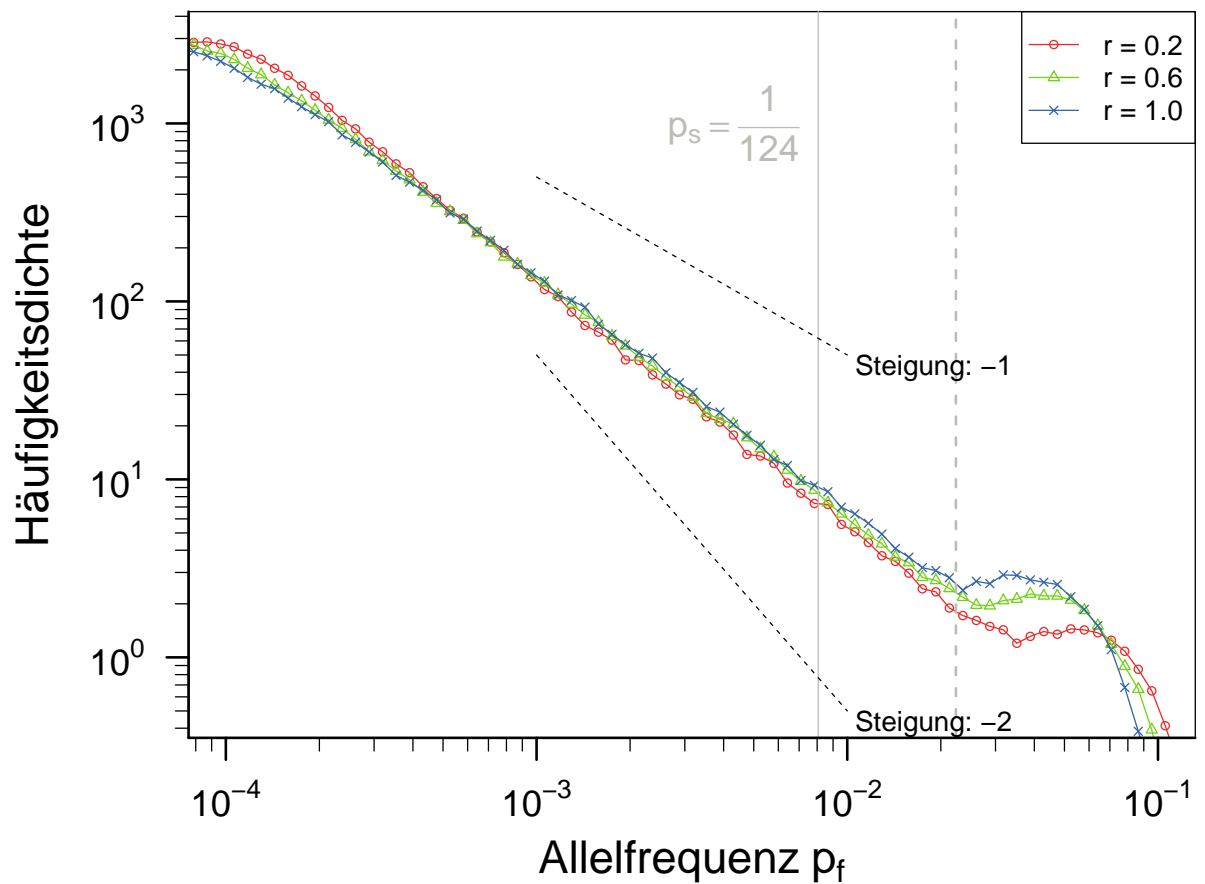


Abbildung B.8.: **Allelfrequenz-Spektren aus Simulationen von Range-Expansions mit verschiedenen Wachstumsraten r .** Im Bereich rechts der gestrichelten vertikalen Linie sind solche Simulationen zu finden, in denen das abgeleitete Allel bei einer längeren Laufzeit noch höhere Frequenzen erreichen könnte. Mit steigender Wachstumsrate erhöht sich die Wahrscheinlichkeit, das abgeleitete Allel nach einer gegebenen Zeit im Bereich hoher Frequenzen zu finden. Ein möglicher Grund ist die verkürzte Zeit, in der das Allel durch genetische Drift aufgrund der kleinen Populationsgröße verloren gehen kann. (TAB. 3.3). Für die verschiedenen Wachstumsraten ist im doppelt-logarithmischen Plot ein Potenzgesetz erkennbar, dessen Exponent im Bereich von $a_k = -1.35$ liegt und mit steigender Wachstumsrate leicht zunimmt (siehe TAB. 3.3 und vergleiche ABB. 3.7). Dieser hohe Exponent deckt sich mit meiner Erwartung für die beobachtete schmalere Front. Die Daten stammen aus Simulationen für Standing-genetic-Variation ($p_s = 0.00806$; durchgezogene vertikale Linie) auf der Großen Welt ($w = 1000$; $h = 1000$; ABB. A.4). Gezeigt sind Ergebnisse aus Simulationen mit einer Kapazitätsgrenze von $K = 100$ und einer Migrationsrate von $m = 0.2$. Die beiden diagonalen Linien zeigen Referenzgeraden mit den Steigungen $a = -1$ und $a = -2$.

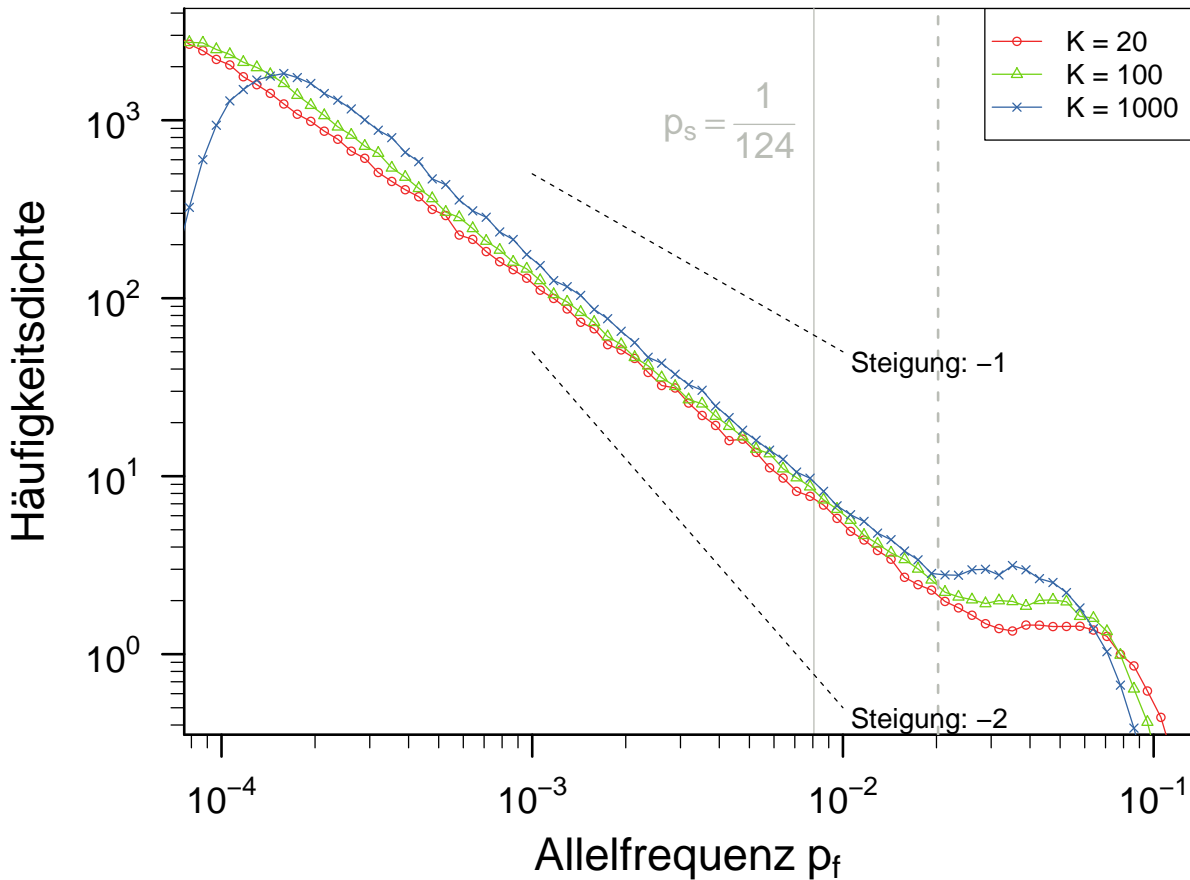


Abbildung B.9.: Allelfrequenz-Spektren aus Simulationen von Range-Expansions mit verschiedenen Kapazitätsgrenzen K . Je höher die Kapazitätsgrenze ist, desto häufiger treten mittlere Allelfrequenzen auf (rechts der gestrichelten vertikalen Linie). Dies deckt sich mit meiner Erwartung, da genetische Drift in kleinen Populationen auf kürzeren Zeitskalen wirkt. Eine höhere Kapazitätsgrenze kann also für eine höhere genetische Vielfalt an der Front verantwortlich sein. Für die verschiedenen Kapazitätsgrenzen ist im doppelt-logarithmischen Plot ein Potenzgesetz erkennbar, dessen Exponent im Bereich von $a_k = -1.4$ liegt (siehe TAB. 3.3 und vergleiche ABB. 3.7). Die Daten stammen aus Simulationen für Standing-genetic-Variation ($p_s = 0.00806$; durchgezogene vertikale Linie) auf der Großen Welt ($w = 1000$; $h = 1000$; ABB. A.4). Gezeigt sind Ergebnisse aus Simulationen mit einer Wachstumsrate von $r = 0.5$ und einer Migrationsrate von $m = 0.2$. Die beiden diagonalen Linien zeigen Referenzgeraden mit den Steigungen $a = -1$ und $a = -2$.

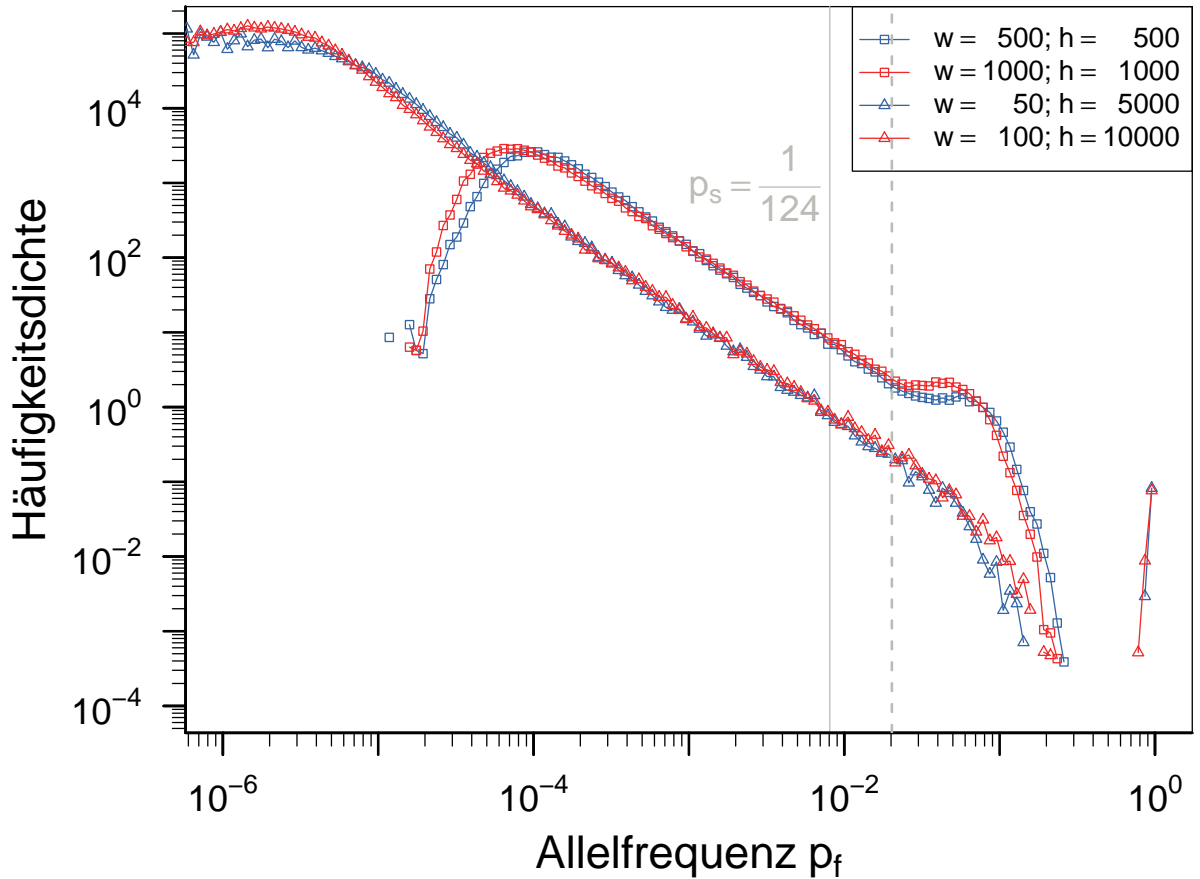


Abbildung B.10.: **Allelfrequenz-Spektren aus Simulationen von Range-Expansions auf Gittern unterschiedlicher Größen und Proportionen.** In Gegensatz zu Simulationen auf den quadratischen Gittern (Vierecke) finden sich in den Simulationen auf der langgezogenen Welt (Dreiecke) solche Mutationen, die bereits nahezu fixiert ($p_f = 1$) sind. Die Zahl der Simulationen in diesem Bereich ist erwartungsgemäß $\simeq \frac{1}{124}\%$ der durchgeführten Simulationen und entspricht damit genau der Fixierungswahrscheinlichkeit der Mutation. Es gibt Frequenzen, die in keiner Simulation auf der langgezogenen Welt (Dreiecke) erreicht werden. Offensichtlich gibt es eine typische Lauflänge, nach der eine Mutation entweder die Front vollständig übernommen oder den Kontakt zur Front verloren hat. Im weiteren Verlauf nähert sich die Frequenz entweder der Fixierung ($p_f = 1$) oder dem Verlust ($p_f = 0$) an. Das Potenzgesetz (siehe z.B. ABB. 3.7) ist sowohl für das langgezogene als auch für das quadratische Gitter messbar. Die Daten stammen aus Simulationen für Standing-genetic-Variation ($p_s = 0.00806$; durchgezogene vertikale Linie) auf den verschiedenen Großen Welten (siehe Legende). Gezeigt sind Ergebnisse aus Simulationen mit einer Wachstumsrate von $r = 0.5$ und einer Migrationsrate von $m = 0.2$.

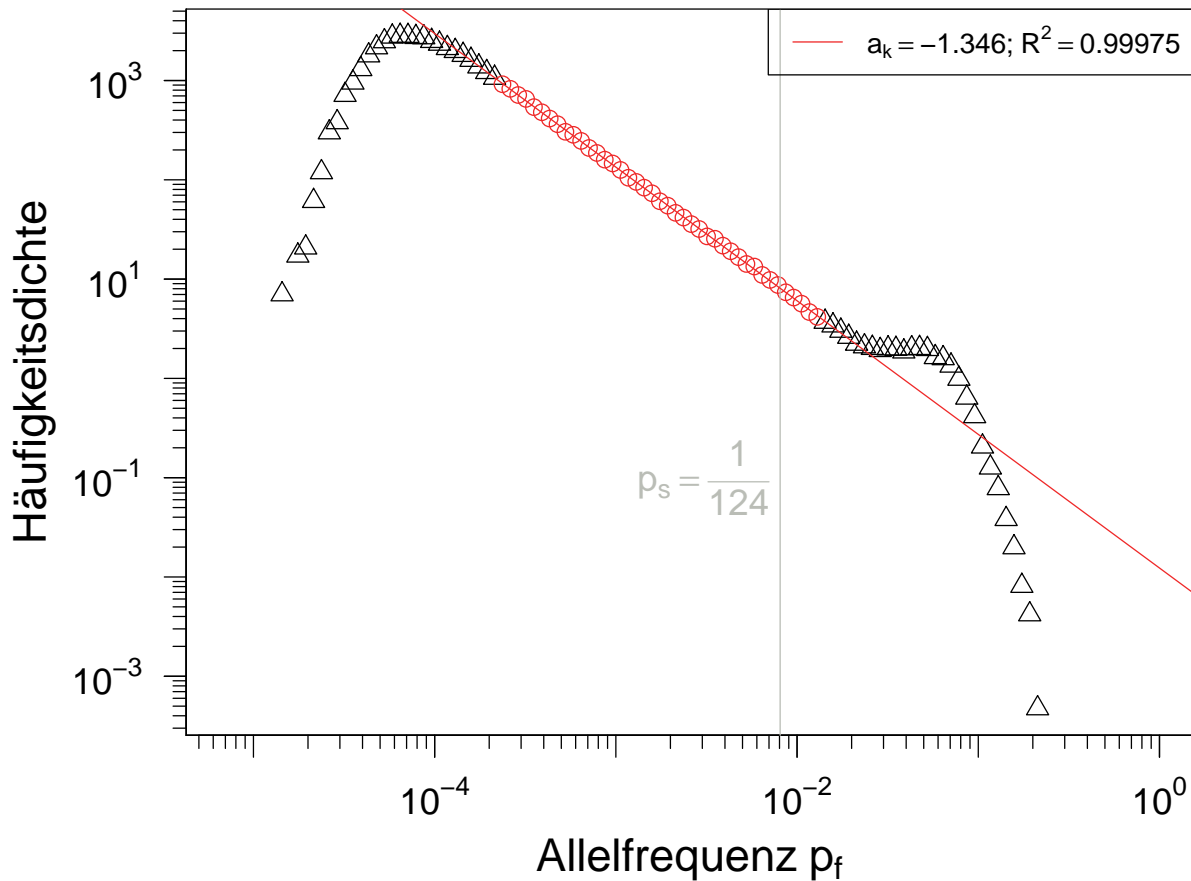


Abbildung B.11.: **Ermittlung des Exponenten a_k des Potenzgesetzes in einem Allelfrequenz-Spektrum klassifizierter Daten.** Dargestellt sind die Häufigkeitsdichten $f(p_f)$, nach der Range-Expansion eine gegebene Frequenz p_f des abgeleiteten Allels zu erhalten. Die Steigung wurde mit Hilfe einer linearen Regression der mit roten Kreisen dargestellten Messwerte ermittelt. Die Steigung a_k und das Bestimmtheitsmaß R^2 sind in der Legende angegeben. Die verwendeten Allelfrequenzen stammen aus Simulationen auf der *Großen Welt* ($p_s = 0.00806$; $K = 100$; $m = 0.2$; $r = 0.5$). Ergebnisse für Simulationen mit anderen Parameterwerten sind in TAB. 3.3 aufgeführt.

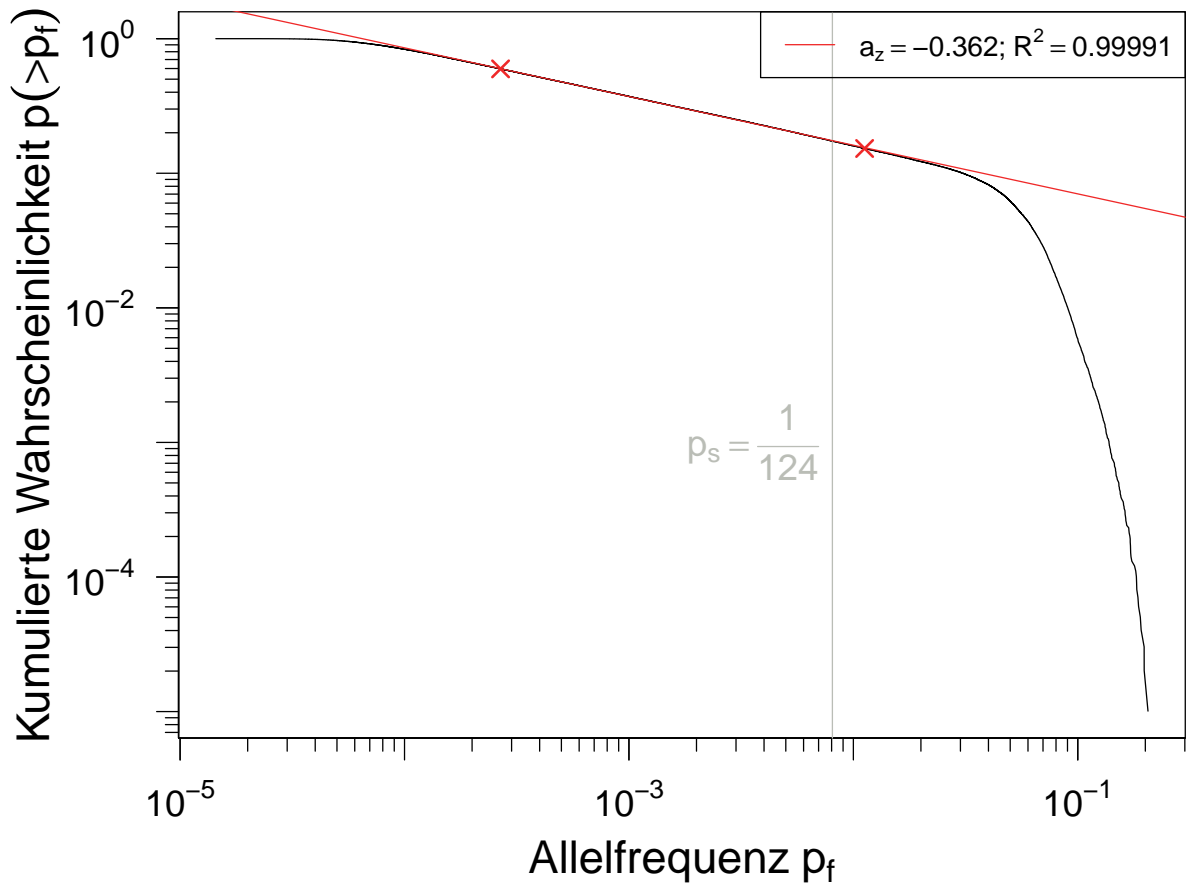


Abbildung B.12.: **Ermittlung des Exponenten a_z des Potenzgesetzes in einem Zipf-Plot.** Dargestellt sind die Wahrscheinlichkeiten $p(> p_f)$, nach der Range-Expansion eine Frequenz des abgeleiteten Allels zu finden, die größer als eine gegebene Frequenz p_f ist. Die Steigung wurde mit Hilfe einer linearen Regression der Messwerte zwischen den beiden markierten Punkten (rote Kreuze) ermittelt. Die Steigung a_z und das Bestimmtheitsmaß R^2 sind in der Legende angegeben. Die verwendeten Allelfrequenzen stammen aus Simulationen auf der *Großen Welt* ($p_s = 0.00806$; $K = 100$; $m = 0.2$; $r = 0.5$). Ergebnisse für Simulationen mit anderen Parameterwerten sind in TAB. 3.3 aufgeführt.

B.2. Theoretische Erwartung für den Exponenten des Potenzgesetzes

Die Exponenten des Potenzgesetzes in den Allelfrequenz-Spektren meiner Simulationen liegen zwischen $a = -1.32$ und $a = -1.46$ (siehe TAB. 3.3). Der Exponent liegt damit zwischen den Erwartungen für eine durchmischte (*well-mixed*) Population stabiler Größe ($a = -1$; Wakeley (2008)) und einer durchmischten, exponentiell wachsenden Population ($a = -2$; Durrett (2013)). Im folgenden Abschnitt präsentiere ich Argumente dafür, dass der erwartete Exponent a des Potenzgesetzes im Bereich von $-1.40 < a < -1.33$ liegt. Diese Ergebnisse basieren auf unpublizierter Arbeit von Matti Gralka, Alexander Anderson und Oskar Hallatschek (Gralka et al., 2014; Anderson and Hallatschek, 2014).

Vereinfacht kann das Problem folgendermaßen betrachtet werden: Mit einer gewissen Wahrscheinlichkeit findet man eine Mutation an der Front. Entweder tritt diese während der Range-Expansion auf oder sie war bereits in der Ursprungspopulation vorhanden. In der Regel verlieren diese Mutationen sehr schnell den Kontakt zur Front und gehen in Folge genetischer Drift verloren (ABB. A.1). Gelegentlich kann es passieren, dass sich eine Mutation an der Front behauptet und eine hohe Frequenz erreicht (ABB. 1.3). In räumlichen Simulationen können auf diese Weise Sektoren entstehen, in denen nur eine genetische Variante vorkommt (ABB. 1.4). Die Fläche A dieser Sektoren ist dabei ein Maß für die Häufigkeit (Frequenz) der Mutation in der Population. Wenn wir vereinfachend annehmen, dass die Front unendlich breit ist, dann wird ein solcher Sektor in jedem Falle irgendwann den Kontakt zur Front verlieren und als *Mutation-Bubble* hinter der Front zurückbleiben (siehe ABB. B.13(B)). Die Frage ist nun, welcher Größenverteilung $p(A)$ diese *Mutation-Bubbles* folgen, denn diese Verteilung ist äquivalent zu den Allelfrequenz-Spektren.

Die Wahrscheinlichkeit $p(> B)$, dass die Mutation-Bubble mindestens eine Breite von B erreicht, ist gleich der Wahrscheinlichkeit $p(> b)$, eine Breite b zu erreichen, mal der bedingten Wahrscheinlichkeit $p(> B | > b)$, eine Breite $> B$ zu erreichen, wenn bereits eine Breite $> b$ erreicht wurde:

$$p(> B) = p(> b) \cdot p(> B | > b) \tag{B.1}$$

Die Sektorgrenzen an der Front kann man als zwei eindimensionale Random-Walker auffassen, die in jedem Zeitschritt entweder einen Schritt nach links oder nach rechts

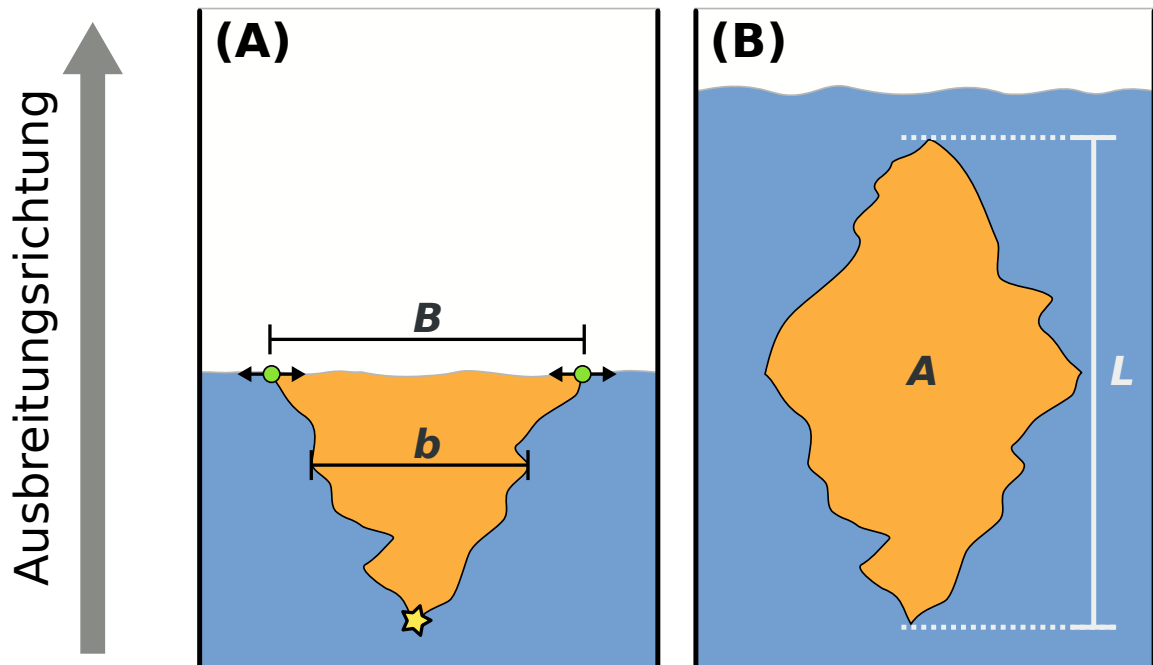


Abbildung B.13.: **Entstehung einer *Mutation-Bubble* während einer Range-Expansion.** (A) Eine Mutation, die an der Front einer expandierenden Population auftritt (Stern), kann durch Gene-Surfing hohe Frequenzen erreichen. Dabei bilden sich Sektoren aus, in denen nur Nachkommen des mutierten Individuums zu finden sind. Die Ränder des Sektors können dabei als Random-Walker (grüne Kreise) betrachtet werden, die in jedem Zeitschritt eine Bewegung nach links oder rechts machen. Treffen sich beide Random-Walker, so verliert die Mutation den Kontakt zur Front (B). Dabei entsteht eine Mutation-Bubble mit der Länge L und der maximalen Breite B . Die Fläche A der Mutation-Bubble ist ein Maß für die Frequenz der Mutation in der Population. Die erwartete Größenverteilung dieser Flächen (und damit die Verteilung der Allelfrequenzen) wird in ABSCHN. B.2 besprochen.

B.2. Theoretische Erwartung für den Exponenten des Potenzgesetzes

gehen (ABB. B.13(A)). Daher entspricht die Sektorbreite einem Random-Walker, der in jedem Zeitschritt zwei Bewegungen macht. Da wir neutrale genetische Variationen betrachten, ist die Wahrscheinlichkeit, dass dieser Random-Walker ausgehend von der Position b eine Position $B = a \cdot b$ (mit $B > b$) erreicht, genau $\frac{1}{a}$. Damit ist:

$$p(> B | > b) = \frac{1}{a} = \frac{b}{b \cdot a} = \frac{b}{B} \quad (\text{B.2})$$

Setzt man nun GL. B.2 in GL. B.1 ein, ergibt sich:

$$p(> B) = \underbrace{p(> b) \cdot b}_C \cdot \frac{1}{B} \sim B^{-1}$$

Da $p(> B)$ nicht von b abhängt, muss für die Gültigkeit der Formel gelten, dass das Produkt aus $p(> b) \cdot b$ für beliebige b konstant ist. Die Wahrscheinlichkeit, dass der Sektor eine Breite $> B$ erreicht, ist also proportional zu $1/B$ (mit der Proportionalitätskonstante C).

Angenommen, ein Sektor hat seit t Generationen Kontakt zur Expansionsfront. Dann ist seine Länge L entlang der Expansionsrichtung bei konstanter Expansionsgeschwindigkeit proportional zu t . Nach dieser Zeit t ist der erwartete Abstand eines Random-Walkers von seiner Startposition, und damit die Breite B des Sektors, proportional zu t^{-z} (Kampen, 2007). Damit gilt:

$$B \sim L^{1/z} \Leftrightarrow L \sim B^z$$

Die Fläche A eines Sektors skaliert mit dem Produkt aus seiner Länge und Breite:

$$A \sim L \cdot B \sim B^z \cdot B^1 \sim B^{z+1} \Leftrightarrow B \sim A^{\frac{1}{z+1}}$$

Für die Wahrscheinlichkeit dafür, dass ein Sektor die Fläche A erreicht, gilt daher:

$$p(> A) \sim p(B(A)) \sim (B(A))^{-1} \sim A^{-\frac{1}{z+1}}$$

Um nun die Wahrscheinlichkeitsdichte $p(A)$ zu erhalten, muss die kumulierte Wahrscheinlichkeitsdichte $p(> A)$ abgeleitet werden:

$$p(A) = p(> A)' = \underbrace{-\frac{1}{z+1}}_{\text{konstant}} \cdot A^{-\frac{1}{z+1}-1} \sim A^{-(\frac{1}{z+1} + \frac{z+1}{z+1})} \sim A^{-\frac{z+2}{z+1}}$$

Der Exponent z der Verteilung $p(A)$ ist dabei abhängig von der Struktur der Expansionsfront. Gerade Fronten findet man bei kurzen Expansionen oder bei Ausbreitungen, deren Geschwindigkeit durch ein sich langsam bewegendes Hindernis (z.B. ein schmelzender Gletscher) begrenzt ist (*boundary limited range expansion*; [Nullmeier and Hallatschek \(2013\)](#)). Für diese geraden Fronten erwartet man eine normale Diffusion der Sektorgrenzen mit $z = 2$ ([Kampen, 2007](#)), wodurch man $p(A) \sim A^{-\frac{4}{3}}$ erhält. Unebene Fronten erwartet man nach längeren Range-Expansions oder durch das Auftreten von Hindernissen. [Hallatschek et al. \(2007\)](#) konnten zeigen, dass sich die Ränder der Sektoren in diesem Falle superdiffusiv bewegen. Die Zahl der Sektoren nimmt dort somit schneller ab als man es bei einem einfachen Random-Walk erwarten würde. Der Exponent z ist in diesem Falle $z = 3/2$, wodurch man für die Verteilung $p(A)$ der Flächengrößen $p(A) \sim A^{-\frac{7}{5}}$ erhält. Die Expansionsfront in meinen Simulationen ist zunächst „gerade“ und wird mit der Zeit „welliger“ (siehe z.B. [ABB. A.2\(A-C\)](#)). Daher erwarte ich, dass die Exponenten der Allelfrequenz-Spektren zwischen diesen beiden Werten liegen, was ich auch in den Simulationsergebnissen finde (siehe [TAB. 3.3](#)).

B.3. Allelfrequenz-Spektren aus echten genetischen Daten

In [ABB. B.14](#) zeige ich die Allelfrequenz-Spektren der abgeleiteten Allele für vier kontinentale Populationen in der doppelt-logarithmischen Darstellung. Darin finde ich im Spektrum für Afrika eine Steigung von $a_{\text{Afrika}} = -1.2$. In den Allelfrequenz-Spektren der drei Populationen (Europa, Asien und Amerika), die vermutlich von Afrika ausgehend neue Teile der Welt besiedelt haben, finde ich hingegen Steigungen die im Bereich von $a = -0.8$ liegen.

In [ABB. B.15](#) zeige ich Allelfrequenz-Spektren von abgeleiteten Allelen, die aller Voraussicht nach während der Range-Expansion entstanden sind (Ongoing-Mutations). Ein mögliches Potenzgesetz im Bereich niedriger Allelfrequenzen kann aufgrund der zu geringen Stichprobengröße nicht zweifelsfrei ermittelt werden.

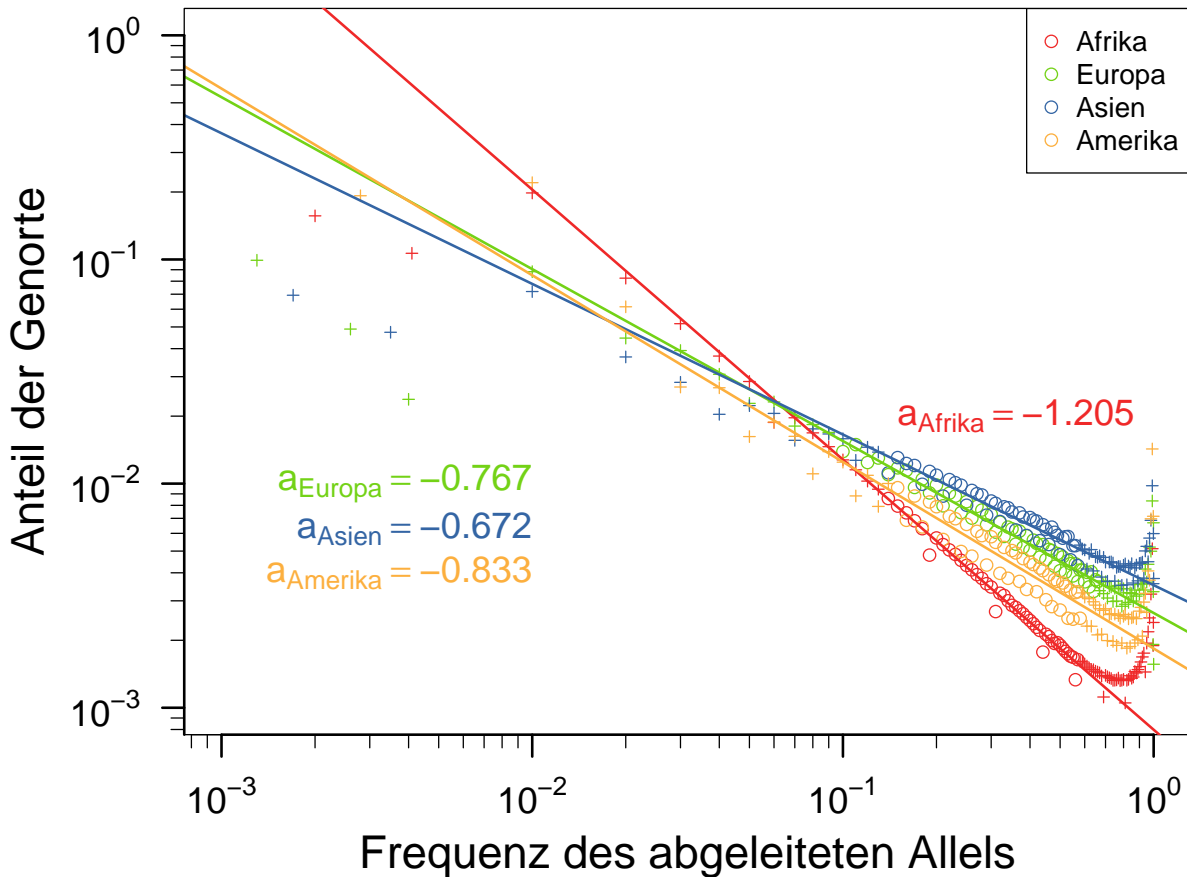


Abbildung B.14.: Allelfrequenz-Spektren für 4 kontinentale Populationen aus Daten des 1000 Genom Projektes. Für die dargestellten Spektren wurden die Frequenzen abgeleiteter Allele aus nicht-codierenden Regionen der Genome verwendet (siehe TEILABSCHN. A.3.2). Es sind nur solche Genorte einbezogen worden, die sowohl in der untersuchten Population als auch in Afrika zu finden sind. Die abgeleiteten Allele an diesen Genorten repräsentieren demnach wahrscheinlich Mutationen, die bereits vor der Ausbreitung des Menschen in der Ursprungspopulation (hier: Afrika) vorhanden waren (Standing-genetic-Variation). Daten, die durch Kreise dargestellt sind, wurden zur Ermittlung der Steigung (lineare Regression) verwendet. Für die Spektren wurden die Frequenzen von abgeleiteten Allelen an $\approx 9.1 \times 10^6$ (Europa), $\approx 7.1 \times 10^6$ (Asien), $\approx 12.5 \times 10^6$ (Amerika) bzw. $\approx 21.5 \times 10^6$ (Afrika) Genorten verwendet. An mehr als der Hälfte der Genorte in afrikanischen Genomen hat das abgeleitete Allel eine Frequenz von $p_f \leq 2\%$. Die nach unten abweichenden Punkte resultieren aus Rundungsfehlern bei der Klassifizierung der Daten.

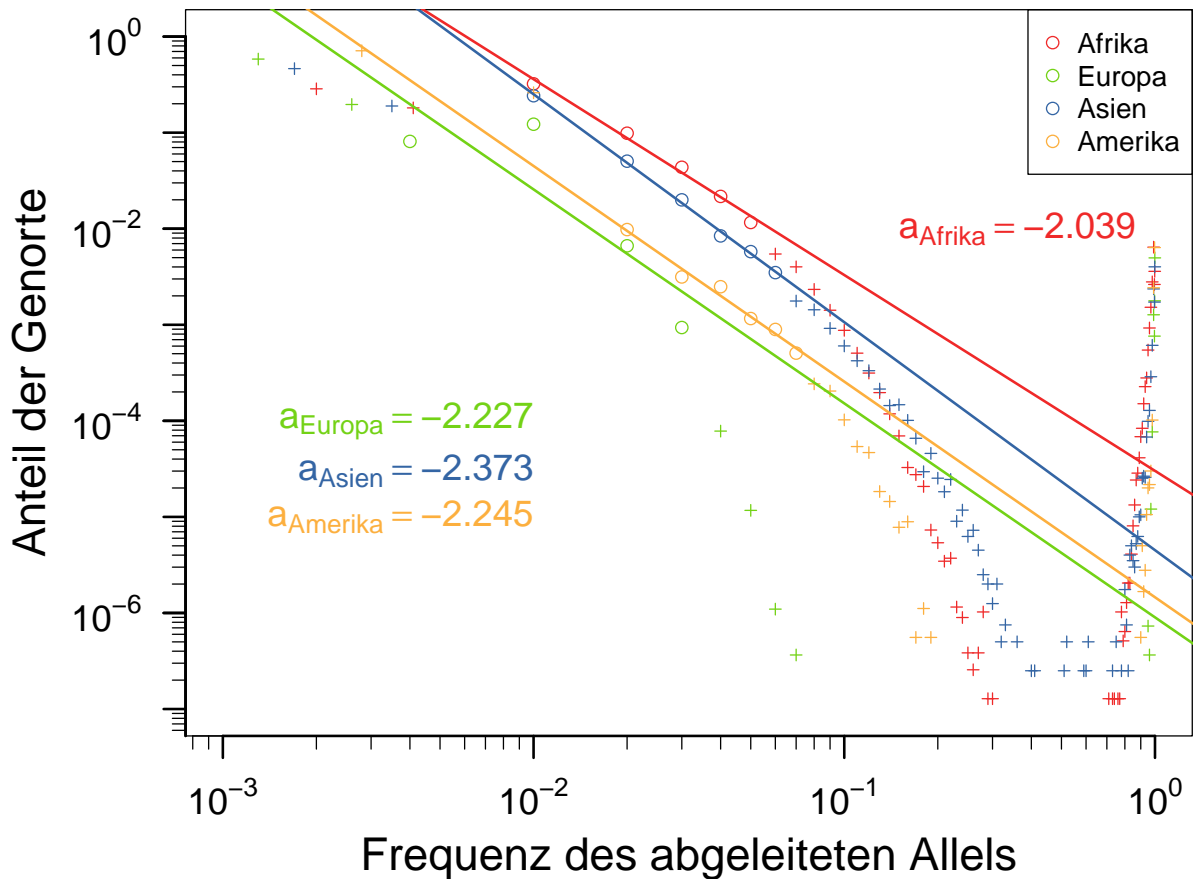


Abbildung B.15.: Allelfrequenz-Spektren für 4 kontinentale Populationen aus Daten des 1000 Genom Projektes. Für die dargestellten Spektren wurden die Frequenzen abgeleiteter Allele aus nicht-codierenden Regionen der Genome verwendet (siehe TEILABSCHN. A.3.2). Im Gegensatz zu ABB. B.14 sind nur Allele betrachtet worden, die ausschließlich in der untersuchten Population zu finden waren. Diese repräsentieren aller Voraussicht nach Mutationen, die nach dem Beginn der Ausbreitung des Menschen entstanden sind (Ongoing-Mutations). Daten, die durch Kreise dargestellt sind, wurden zur Ermittlung der Steigung (lineare Regression) verwendet. Die Steigungen sind dabei nur als Orientierungshilfe zu verstehen, da die Auflösung für die Bestimmung eines eventuellen Potenzgesetzes zu gering ist (weniger als eine Größenordnung). Für die Spektren wurden die Frequenzen von abgeleiteten Allelen an $\approx 2.7 \times 10^6$ (Europa), $\approx 4.0 \times 10^6$ (Asien), $\approx 1.8 \times 10^6$ (Amerika) bzw. $\approx 7.8 \times 10^6$ (Afrika) Genorten verwendet.

C. Anhang (Bildquellen)

Die Lizenzen aller Bilder, die ich in dieser Arbeit verwendet habe, erlauben die Verwendung für nicht kommerzielle Zwecke, sofern auf den Rechteinhaber verwiesen wird.

- ABB. 1.1(A): Das Bild zeigt eine Rekonstruktion des 'Mannes vom Hauslabjoch' (auch bekannt als 'Ötzi'); © Südtiroler Archäologiemuseum/A. Oxsenreiter. (archaeologiemuseum.it/en/photo-archive; 11.12.2013)
- ABB. 1.1(B): Gemeiner Grashüpfer (*Chorthippus parallelus*); Weibchen. (en.wikipedia.org/wiki/File:Chorthippus_parallelus_femelle_1.jpg; 11.12.2013)
- ABB. 1.1(C): Foto einer weiblichen Aga-Kröte (*Rhinella marina*); © Australian Museum/G. Millen. (australianmuseum.net.au/image/Cane-Toad-front-view/; 11.12.2013)
- ABB. 1.1(D): Foto des Gewöhnlichen Blutweiderich *Lythrum salicaria*, aufgenommen in Schweden von Jonas Bergsten. (commons.wikimedia.org/wiki/File:Lythrum_salicaria_20060712_002.jpg; 11.12.2013)
- ABB. 1.1(E): Foto einer Zebramuschel (*Dreissena polymorpha*). © US Geological Survey. [2013]. Nonindigenous Aquatic Species Database. Gainesville, Florida. (nas.er.usgs.gov/taxgroup/mollusks/zebramussel/; 11.12.2013)
- ABB. 1.1(F): Fluoreszenzmikroskopische Aufnahme einer Bakterienkolonie, die zwei Tage auf einer Agarose-Platte gewachsen ist. Den Ursprung der Population bildete ein 1:1-Gemisch zweier *Escherichia coli* Stämme, von denen einer das Gelb fluoreszierende Protein (*YFP*) und der andere das Cyan fluoreszierende Protein (*CFP*) produziert. Die Produktion dieser Proteine hat keinen Einfluss auf die Fitness der Bakterien. Das Bild wurde mir freundlicherweise von Fabian Stiewe zur Verfügung gestellt.
- Die Abbildungen ABB. 1.2 und ABB. A.7 zeigen eine Weltkarte in der Robinson-Projektion. Als Vorlage diente mir hierbei: (commons.wikimedia.org/wiki/File:WorldMap.svg; 16.12.2013)

C. Anhang (Bildquellen)

- Für die Abbildungen ABB. A.1, ABB. 2.3 und ABB. 2.4 habe ich folgendes Bild als Vorlage verwendet:
(commons.wikimedia.org/wiki/File:Random_sampling_genetic_drift.svg; 11.02.2014)

Lebenslauf

Persönliche Daten

Name: Sven Boekhoff
Geburtsdatum: 13.07.1982
Geburtsort: Westerstede, Deutschland
Familienstand: Verheiratet
Staatsangehörigkeit: Deutsch

Ausbildung

- 2009–2014** Doktorand am Max-Planck-Institut für Dynamik und Selbstorganisation, Göttingen. Projekt: *Quantifying genetic footprints of natural selection and demographic changes* (Betreut durch Dr. Oskar Hallatschek, Prof. Dr. Dirk Semmann und Prof. Dr. Kerstin Wiegand)
- 2009** Master Abschluss (Biological Sciences) an der Universität Konstanz. Projekt: *Evolutionary genomic analysis of cichlid fishes based on large scale sequence data* (Betreut durch Prof. Dr. Axel Meyer und Dr. Shigehiro Kuraku)
- 2006** Bachelor Abschluss (Biological Sciences) an der Universität Konstanz, in Zusammenarbeit mit dem IGB Berlin. Projekt: *Succession and diversity of eubacteria and fungi during leaf litter decomposition in the littoral zone of Lake Konstanz* (Betreut durch Prof. Dr. Karl-Otto Rothhaupt (Univ. Konstanz), PD Dr. Hans-Peter Grossart (IGB Berlin) und PD Dr. Karl Matthias Wantzen (Univ. Konstanz))
- 2001** Abitur an der Europaschule Gymnasium Westerstede, Westerstede

Danksagung

Viele Menschen haben die Zeit meiner Doktorarbeit entscheidend geprägt. Einige möchte ich an dieser Stelle erwähnen.

Dr. Oskar Hallatschek für die Möglichkeit, das Thema in seiner Arbeitsgruppe zu bearbeiten sowie die zahlreichen Ideen.

Prof. Dr. Dirk Semmann für die Motivation sowie die Integration in seine Gruppe und das CRC.

Prof. Dr. Kerstin Wiegand für die fachliche und menschliche Unterstützung.

Allen Kommissionsmitgliedern für ihre Bereitschaft, diese Arbeit zu bewerten.

Allen aktuellen und früheren Mitgliedern der Gruppe „Biologische Physik und Evolutionäre Dynamik“. Insbesondere danke ich Lukas, Jens, Erik, Fabian und Jörn für fachliche Diskussionen und hilfreiche Kommentare zu den Versionen dieser Arbeit.

Katrin Meyer für hilfreiche Gespräche.

Aus dem wissenschaftlichen Umfeld: Peter Bednarik, Bernhard Altaner, Artur Wachtel, Katrin Fehl, Barbara Kutz. Vielen Mitgliedern des CRC (u.a. Michael Belz, Karo Lohse und Gisela Kopp) und der „Young Modellers in Ecology“.

Aus meinem privaten Umfeld: Eike Grüßing, Swetlana und Alexej Konkov, Marcel Sinemus, Josefine Wahl und Rüdiger Link.

Ganz besonderer Dank gilt meinen Eltern, die mich in all meinen Entscheidungen unterstützt und bestärkt haben.

... und natürlich Irina, Anita und Alena für all die schönen Dinge, die man nicht in Worte fassen kann.