

Ihre Namen: \_\_\_\_\_

## Übung A: Populationsgenetik

### Teil 1: Mendelsche Vererbung

1) Schwarze Fellfarbe bei Pferden ist weitgehend bestimmt durch ein rezessives Allel des *A*-Locus. *AA*- bzw. *Aa*-Pferde haben nichtschwarzes Fell, z.B. braun, während *aa*-Pferde schwarzes Fell aufweisen. Vor ein paar Jahren fragte ein Leser der Internet-Newsgroup rec.equestrian, warum es relativ wenige schwarze Pferde in der Araber-Rassengruppe gibt. Eine Antwort war: „Mehr Araber sind braun oder grau, weil diese Farben dominant sind. Schwarz ist eine seltene Farbe, weil das schwarze Allel rezessiv ist. Schwarze Araber sind im Laufe der Zeit immer weniger geworden.“ Was ist falsch an dieser Erklärung?

2) In einer Aquariums-Population von 125 Buntbarschen beobachten Sie drei verschiedene Farbausprägungen: 60 Fische sind leuchtend gelb, 30 Fische zeigen eine grünliche Färbung und 35 Fische sind dunkelblau. Die Farbe dieser Fische wird vom *B*-Locus bestimmt, an dem es zwei Allele gibt: *B* und *b*. Individuen mit dem *BB*-Genotyp sind gelb, heterozygote *Bb*-Individuen sind grün und *bb*-Individuen sind blau.

a) Entspricht dieser Locus dem Hardy-Weinberg-Gleichgewicht? Führen Sie hierzu den in der Vorlesung besprochenen Chi-Quadrat-Test durch. Als Chi-Quadrat Wert benutzen wir die Irrtumswahrscheinlichkeit 0,05 und 1 Freiheitsgrad, d.h. die Nullhypothese wird abgelehnt, wenn der Chi-Quadrat-Wert größer ist als 3,84.

b) Was könnte der Grund für die beobachteten Abweichungen sein? Bedenken Sie, dass bei vielen Buntbarscharten die Farbe eine wichtige Rolle bei der Partnerwahl spielt. Welche Voraussetzung einer idealen Hardy-Weinberg-Population ist in diesem Fall nicht erfüllt?

## Teil 2: Genetische Drift

Im zweiten Teil der heutigen Übungen wollen wir untersuchen, wie sich Allelfrequenzen in Populationen im Laufe der Zeit ändern. In endlich großen Populationen findet genetische Drift statt. Um den Effekt von genetischer Drift auf Variabilität innerhalb einer Population besser zu verstehen, benutzen wir das aus der Vorlesung bekannte Wright-Fisher-Modell.

### **Manuelle Simulation eines Wright-Fisher-Modells durch „Würfeln“**

In dieser Übung wollen wir ein Wright-Fisher-Modell "auswürfeln". Die Zahl der Seiten des Würfels entspricht dabei der Anzahl an Individuen in einer Population. Damit lässt sich die Zufallspaarung wie sie in diesem Modell von einer Generation zur Nächsten angenommen wird simulieren.

Anstatt tatsächliche Würfel zu benutzen, verwenden wir das Online-Tool

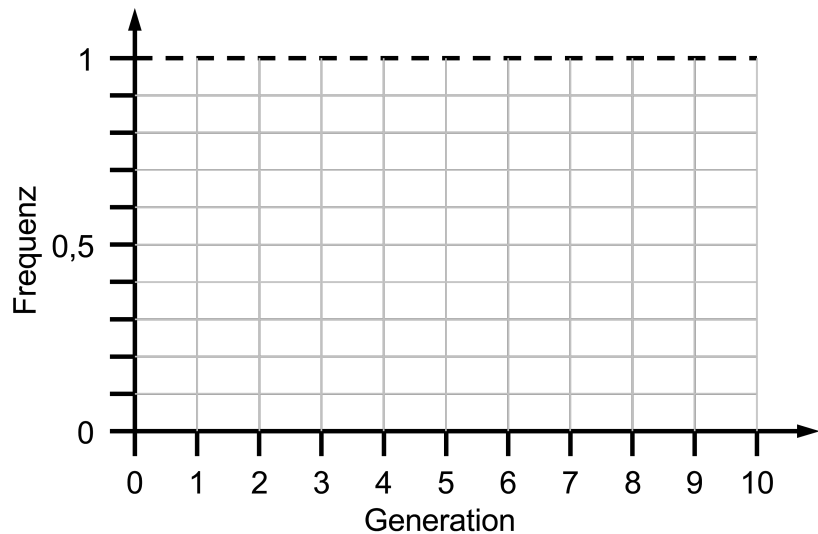
<http://www.roll-dice-online.com>.

Simulieren wir hiermit nun eine Wright-Fisher-Population. Betrachten Sie die untere Tabelle. Jede Zeile stellt eine Generation Ihrer Population dar. Die Spaltenzahl entspricht der Anzahl an Individuen in Ihrer Population, sowie der Anzahl an Seiten des Würfels den Sie benutzen.

Individuum:	1	2	3	4	5	6	7	8	9	10
Generation 10										
Generation 9										
Generation 8										
Generation 7										
Generation 6										
Generation 5										
Generation 4										
Generation 3										
Generation 2										
Generation 1										
Generation 0										

**3a)** Beginnen Sie in der untersten Zeile (Generation 0). Sorgen Sie zunächst für „genetische Variabilität“ indem sie die Hälfte aller Individuen zufällig schraffieren. Sie erzeugen dadurch zwei Allele (schraffiert und weiß). Beide Allele starten also mit einer Frequenz von 50%. Würfeln Sie nun die Nachkommen in den folgenden Generationen nach dem Wright-Fisher-Prinzip aus. Stellen Sie dazu im Online-Tool *Number of sides = 10, Number of dice to roll = 10* und *number of rolls = 10* und würfeln Sie (*roll dice*).

Sie erhalten nun 10 Zeilen mit 10 Würfelergebnissen. Jede Zeile repräsentiert eine Generation (Generationen 1 - 10) in Ihrer Wright-Fisher-Population. Die 10 Würfelergebnisse bestimmen die Eltern jedes der 10 Individuen in dieser Generation. Fangen Sie mit der untersten Reihe an. Hier wurden die Eltern aller Individuen in Generation 1 ausgewürfelt. Die erste Zahl bestimmt den Vorfahren von Individuum 1. Sollte hier z.B. eine 4 stehen, dann bedeutet dies, dass Individuum 4 in Generation 0 der Vorfahre war. Dieser Vorfahre vererbt sein Allel (schraffiert oder weiß) an seinen Nachkommen weiter. Die zweite Zahl steht für den Vorfahren von Individuum 2, usw. Somit kann die Eltern-Nachkommenschaft zwischen den Generationen 0 und 1 ermittelt werden und damit auch die Allele die vererbt wurden. Ermitteln Sie so die Allele, welche die 10 Individuen in Generation 1 tragen. Betrachten Sie nun die zweite Zeile von unten. Diese definiert die Eltern-Nachkommenschaft zwischen den Generationen 1 und 2. Jede Zahl bestimmt den Vorfahren eines Individuums aus Generation 2 (der aus Generation 1 stammt) und Sie können dadurch ermitteln welche Allele in Generation 2 weitervererbt wurden. Bestimmen Sie nach diesem Schema den Erbgang über 10 Generationen und füllen Sie die Tabelle mit den jeweiligen Allelen in jeder Generation aus.



**3b)** Obwohl Sie in Ihrer Wright-Fisher-Population in Generation 0 mit einer Frequenz von 50% für das schraffierte Allel angefangen haben, hat sich diese Frequenz (höchstwahrscheinlich) im Laufe der Zeit verändert. Diese Änderung in der Allelfrequenz wollen wir graphisch festhalten. Zeichnen sie die Änderung, die Sie in Ihrer Population sehen, im Graphen rechts ein.

### Wright-Fisher-Simulationen mit *Populus*

Im nächsten Übungsteil werden wir nun das Computerprogramm *Populus* anwenden, um genauere Einblicke in die Konsequenzen von genetischer Drift zu erhalten.

Öffnen Sie das Programm *Populus*. Wählen Sie unter *Model Mendelian Genetics > Genetic Drift* und dort *Monte Carlo*. Diese Funktion von *Populus* führt Wright-Fischer-Simulationen durch, ganz ähnlich, wie Sie diese im vorherigen Übungsteil kennengelernt haben. Es wird das Schicksal von zwei Allelen an einem Locus über einen gegebenen Zeitraum simuliert. Klassischerweise werden die beiden Allele mit *A* und *a* bezeichnet und die Frequenz von *A* mit *p* angegeben. Bei diesen Simulationen findet keine Selektion statt! D.h. die unterschiedlichen Allele haben keinen Einfluss auf die Fitness des Organismus.

Machen Sie sich erst einmal mit den unterschiedlichen Einstellungsmöglichkeiten vertraut. Setzen Sie zunächst *Number of Loci = 1, Population Size (N) = 10* und belassen Sie die anderen

Einstellungen. Starten Sie eine Simulation indem sie auf *View* klicken oder einfach auf der Tastatur „Enter“ drücken.

**4a)** Interpretieren Sie den Output des Programms. Was ist auf der X- und was auf der Y-Achse aufgetragen? Wo sind die Gemeinsamkeiten zum „Output“ der Simulationen, die Sie in Übung 2 manuell durchgeführt haben?

**4b)** Mit dem Parameter *Population Size (N)* kann die Anzahl der Individuen eingestellt werden. Allerdings simuliert POPULUS hier diploide Individuen. Was bedeutet das für die tatsächliche Anzahl an Genkopien in der Population?

Setzen Sie nun *Number of Loci = 10* und klicken Sie bei *Runtime* auf *Other* und setzen Sie *Generations = 50*. Hier werden nun 10 Loci unabhängig voneinander für 50 Generationen simuliert (d.h. es finden eigentlich 10 getrennte Simulationen statt) und das Ergebnis wird in einem einzigen Graphen ausgegeben wobei jeder Locus in einer anderen Farbe dargestellt wird. Lassen Sie die Simulationen einige Male laufen um ein Gefühl dafür zu bekommen.

**5a)** Wir wollen nun den Einfluss der Populationsgröße auf die Dynamik der Allele untersuchen.

Vergleichen Sie den Effekt der Populationsgröße  $N = 10$ ,  $N = 20$  und  $N = 50$ . Lassen Sie die Simulation 10-mal für jede Populationsgrößen laufen und notieren Sie für jeden Durchgang wie viele der 10 einzelnen Loci innerhalb der simulierten 50 Generationen eines der beiden Allele fixiert haben, d.h. die Frequenz des Allels eines Locus erreicht  $p = 0$  oder  $p = 1$ . Berechnen Sie zudem den Mittelwert über alle Durchgänge. Benutzen Sie hierzu die Tabelle rechts.

Durchgang	Populationsgröße		
	$N = 10$	$N = 20$	$N = 50$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
Mittelwert			

**5b)** Tragen Sie Ihren Mittelwert in die Excel-Tabelle des Dozenten ein. Notieren Sie den Mittelwert über alle Gruppen hier:  $N=10$  :                       $N=20$ :                       $N=50$ :

**5c)** Der Verlust eines Allels ist ein extremer Ausdruck der genetischen Drift. Hat dabei die Populationsgröße einen Effekt auf den möglichen Verlust eines Allels?

**6a)** Bis jetzt war uns bei Fixierungsereignissen egal ob  $A$  ( $p = 1$ ) oder  $a$  ( $p = 0$ ) fixiert wurde. Betrachten wir dies nun etwas differenzierter. Setzen Sie  $N = 10$  und  $Generations = 1000$ . Da die Zeitspanne so groß ist, sollten bei diesen Simulationen innerhalb der simulierten Zeit nun alle Loci fixieren (d.h. das Fenster mit dem Hinweis „All loci have fixed by  $t = x$  generations“ sollte erscheinen). Führen Sie 10 Simulationsdurchgänge durch und notieren Sie für jeden Durchgang wie viele der 10 Loci das  $A$ -Allel fixiert haben, nicht das  $a$  Allel (d.h. wenn  $p = 1$  erreicht). In anderen Worten: schreiben Sie auf, wie viele Loci den oberen Rand erreicht haben. Wiederholen Sie die Simulationen, ändern Sie aber die Startfrequenz von  $A$  (*initial frequency*) auf 0,2 sowie auf 0,7. Sammeln Sie die Ergebnisse in der Tabelle rechts und berechnen Sie die jeweiligen Mittelwerte.

Durchgang	Startfrequenz von A		
	$p = 0,2$	$p = 0,5$	$p = 0,7$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
Mittelwert			

**6b)** Tragen Sie Ihre Mittelwerte in die Excel-Tabelle des Dozenten ein. Notieren Sie die Mittelwerte über alle Gruppen hier:  $p=0,2$ :                       $p=0,5$ :                       $p=0,7$ :

**6c)** Welcher Zusammenhang könnte zwischen der Startfrequenz und der Fixierungswahrscheinlichkeit bestehen?

### Teil 3: Gerichtete Selektion

Die Variabilität die wir bis jetzt betrachtet haben war selektiv neutral, d.h. keines der beiden Allele hatte einen Vorteil. Dies wollen wir nun ändern und mit Hilfe von Simulationen uns mit den Konsequenzen von Selektion auf die Dynamik von Allelen innerhalb einer Population beschäftigen. Dies ist die Grundlage für Evolution durch natürliche Selektion wie Darwin sie bereits vor über 150 Jahren beschrieben hat: Eine Variante hat innerhalb der Population (oder Art) einen Vorteil und wird sich im Laufe der Zeit durchsetzen.

Wählen Sie in POPULUS das Model *Natural Selection -> Selection on a Diallelic Autosomal Locus*. Hier wird der Effekt von Selektion modelliert und zwar genau nach den Prinzipien, die Sie bereits aus der Vorlesung kennen. Es wird von einem Locus mit zwei Allelen und drei Genotypen ausgegangen. Die Population wird als unendlich groß angenommen. Das bedeutet auch, dass es keine Drift gibt. Tatsächlich werden die Kurven in diesem Versuchsteil nicht simuliert, sondern einfach vom Computer berechnet. (Man sagt: es ist ein deterministisches Modell.)

**7a)** Betrachten Sie zunächst die standardmäßig vorgegeben Einstellungen für *Fitness/Selection Coeffs*. Was bedeuten die Parameter  $w_{AA}$ ,  $w_{Aa}$  und  $w_{aa}$ ?

**7b)** Können Sie allein aus den gegebenen Zahlen ablesen, welches der beiden Allele einen Selektionsvorteil besitzt?

**7c)** Lassen Sie die Simulation mit den vorgegebenen Einstellungen laufen, indem Sie auf *View* klicken. Was wird im ausgegebenen Graph dargestellt? Entspricht das Ergebnis Ihren Erwartungen?

**8a)** Simulieren wir nun eine Population in dem das *A* Allel dominant ist und einen Fitnessvorteil von 50% hat. Wie müssen Sie in diesem Fall die Werte von  $w_{AA}$ ,  $w_{Aa}$  und  $w_{aa}$  ändern?

**8b)** Ändern Sie nun den Fitnessvorteil des *A* Allels und betrachten Sie Situationen mit unterschiedlichen Fitnessvorteilen (5%, 10% und 20%). Wie verändert dies die Form dieses Graphen?

**9a)** Im Folgenden wollen wir untersuchen, wie sich im Vergleich zu oben Selektion unterscheidet, wenn die vorteilhafte Mutation *A* rezessiv ist. Simulieren Sie eine rezessive Mutation mit einem Fitnessvorteil von 20%. Welche Fitnesswerte müssen Sie gegenüber einer dominanten Mutation von 20% ändern? Betrachten Sie die Dynamik der Allele vor allem in der Anfangsphase. Haben Sie eine Erklärung für die Unterschiede zwischen der dominanten und der rezessiven Situation?

**9b)** Betrachten Sie nun die Phase, in der das positiv selektierte Allel kurz vor der Fixierung steht (d.h. die Frequenz nähert sich 1). Vergleichen Sie die dominante und rezessive Simulation, indem Sie eine zweite Simulation über Model *Natural Selection* -> *Selection on a Diallelic Autosomal Locus* starten. Aktivieren Sie im Ausgabefenster durch mehrmaliges Klicken auf *Options* -> *Finer Grid* ein Gitternetz, was das Ablesen erleichtert. Außerdem können in das Ausgabefenster hinein zoomen, wenn Sie mit der Maus einen Kasten ziehen. Mit *Options* -> *Reset Graph* können Sie das Ausgabefenster wieder zurücksetzen. Bei welchen der beiden Situationen wird das selektierte Allel schneller fixiert? Hätten Sie dies erwartet? Fällt Ihnen eine mögliche Erklärung dafür ein?

#### Teil 4: Zusammenspiel von Selektion und Drift

Im letzten Teil der Übung sind wir vom Fall ausgegangen, dass sich eine positiv selektierte Mutation immer durchsetzen wird (über kurz oder lang). Dabei haben wir den Effekt von genetischer Drift völlig außer acht gelassen. Mit dem nächsten Modell wollen wir nun Drift zusätzlich zu Selektion betrachten.

**10)** Wählen Sie in POPULUS das Modell *Mendelian Genetics* -> *Drift and Selection*. Machen Sie sich mit den verschiedenen Einstellungsmöglichkeiten vertraut. Welcher Parameter in diesem Modell ist für die Drift zuständig?

**11a)** Simulieren Sie nun in diesem Modell eine dominante Mutation mit Fitnessvorteil 20%, Startfrequenz 10% und setzen sie *Population Size (N)* = 500. Vergleichen Sie die Dynamik der Allele hier mit einer Simulation in einer unendlich großen Population (benutzen Sie dazu wieder das Model *Natural Selection* -> *Selection on a Diallelic Autosomal Locus* und vergleichen Sie beide Ausgabefenster untereinander). Worin unterscheiden sie sich? Was ist für diesen Effekt verantwortlich?

**11b)** Wiederholen Sie Ihre Simulationen, aber verringern sie die Populationsgröße (auf 100, 50 und 10). Lassen Sie die Simulationen mehrmals mit den verringerten Populationsgrößen laufen. Setzt sich das positiv selektierte Allel immer durch? Erklären Sie Ihre Beobachtungen.

**12a)** Wir wollen den in Übung 10 beobachteten Effekt etwas genauer quantifizieren. Setzen Sie die Fitnesswerte auf  $w_{AA} = 1,0$ ,  $w_{Aa} = 0,9$  und  $w_{aa} = 0,9$ . Setzen Sie die Startfrequenz  $p = 0,5$  und *Runtime (generations)* auf 500 Generationen. Lassen Sie nun für die Populationsgrößen  $N = 10$ , 20 und 50 jeweils 40 Simulationen laufen, und verfolgen Sie dabei, wie häufig das A-Allel fixiert ( $p = 1$ ) und wie häufig das a-Allel fixiert ( $p = 0$ ). Notieren Sie Ihre Ergebnisse in der Tabelle und geben Sie diese dann in die Excel-Tabelle des Dozenten ein um den Durchschnitt aller Gruppen zu ermitteln.

	$N = 10$		$N = 20$		$N = 50$	
	$p = 1$	$p = 0$	$p = 1$	$p = 0$	$p = 1$	$p = 0$
<b>Gruppe:</b>						
<b>Durchschnitt aller Gruppen:</b>						

**12b)** Welchen Einfluss hat die Populationsgröße auf die Fixierung des A-Allels in Ihren Simulationen? Was können Sie darüber sagen, wie sich genetische Drift auf den Ausgang der natürlichen Selektion auswirkt?

**13a)** Verdoppeln Sie nun den Selektionskoeffizienten der Mutation aus Übung 12 ( $w_{AA} = 1,0$ ,  $w_{Aa} = 0,8$  und  $w_{aa} = 0,8$ ) und wiederholen Sie die Versuchsreihe. Notieren Sie Ihre Ergebnisse hier und geben Sie diese dann in die Excel-Tabelle des Dozenten ein um den Durchschnitt aller Gruppen zu ermitteln.

	$N = 10$		$N = 20$		$N = 50$	
	$p = 1$	$p = 0$	$p = 1$	$p = 0$	$p = 1$	$p = 0$
<b>Gruppe:</b>						
<b>Durchschnitt aller Gruppen:</b>						

**13b)** Vergleichen Sie Ihre Ergebnisse mit den Werten aus Übung 11. Können Sie einen Zusammenhang zwischen Populationsgröße, Selektionskoeffizienten und der Fixierungswahrscheinlichkeit  $\pi(p)$  des positiv selektierten Allels erkennen?