

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

**Evolutionsbiologie 2, WS2016/2017: Bioinformatik - Übung 3**

**Populationsgenetische Analyse von vier unbekanntem Loci aus drei *Drosophila melanogaster* Populationen (Zimbabwe, Afrika; Niederlande, Europa; Malaysia, Asien)**

Summary statistics

Locus		n	bp	S	$\theta_w$ (per site)	$\pi$ (per site)	Tajima's <i>D</i>	<i>K</i>
Afrika	55							
	126							
	530							
	1435							
Europa	55							
	126							
	530							
	1435							
Asien	55							
	126							
	530							
	1435							

Genetische Differenzierung zwischen den Populationen

$F_{ST}$	Afrika - Europa	Afrika - Asien	Europa - Asien
55			
126			
530			
1435			

1) Sie haben insgesamt vier Loci aus drei *D. melanogaster* Populationen analysiert. Diese wurden ursprünglich sequenziert, weil man annahm, dass sie neutral evolvieren. Es stellte sich jedoch heraus, dass einer der Loci vermutlich in einem genomischen Bereich liegt in dem positive Selektion aufgrund von lokaler Adaptation herrscht. Können Sie aufgrund der Werte erkennen welcher der vier Loci nicht neutral evolviert, also von den anderen abweicht und somit möglicherweise unter Selektion steht? Begründen Sie ihre Vermutung. (Hinweis: wir werden den selektierten Locus für die nächsten Analyseschritte ignorieren. Wenn Sie unsicher sind ob Sie den richtigen Locus identifiziert haben, fragen Sie einen Tutor).

2) Zunächst werden Sie den demographischen Hintergrund der Populationen näher betrachten. Berechnen Sie dazu die Durchschnittswerte der drei neutral evolvierenden Loci für die angegebenen Statistiken.

	$\theta_w$ (per site)	$\pi$ (per site)	Tajima's <i>D</i>	<i>K</i>
Ø Afrika				
Ø Europa				
Ø Asien				

a) Basierend auf den Statistiken, was für demographische Szenarien sind für die drei Populationen jeweils die wahrscheinlichsten? Welche Populationen sind alt/jung, bei welchen Populationen hat Populationswachstum/Populationsrückgang stattgefunden?

b) Vergleichen Sie nun die durchschnittliche genetische Differenzierung der drei neutralen Loci zwischen den drei Populationen. Welche Populationen sind am stärksten voneinander differenziert, welche sind sich am ähnlichsten?

	<b>Afrika - Europa</b>	<b>Afrika - Asien</b>	<b>Europa - Asien</b>
$\bar{\theta} F_{ST}$			

c) Mit Hilfe Ihrer Erkenntnisse aus 2a und 2b entwerfen sie die weltweite demographische Geschichte von *D. melanogaster*. Welche Population auf welchem Kontinent ist die ursprünglichste und somit nahe am Artenstehungszentrum? In welcher Reihenfolge wurden die anderen Kontinente besiedelt?

d) Die Divergenz zu einer Außengruppe ( $K$ ) an einem neutralen Locus liefert einen Hinweis über die Mutationsrate die an diesem Locus herrscht. An welchem Ihrer neutralen Loci würde demnach eine besonders niedrige Mutationsrate herrschen?

3a) Mit Hilfe von neutraler Variabilität (z.B.  $\theta_w$  an neutralen Loci) lässt sich die effektive Populationsgröße einer Population ( $N_e$ ) berechnen. Wie lautet die Formel hierzu? (Hinweis: Vorlesung)

b) Berechnen Sie  $N_e$  für die drei Populationen. Benutzen Sie den Durchschnittswert von  $\theta_w$  der neutralen Loci und nehmen Sie eine durchschnittliche Mutationsrate von  $\mu = 1,5 \times 10^{-9}$  an.

c) Entsprechen die Unterschiede in  $N_e$  zwischen den Populationen zwingend einem tatsächlichen Unterschied an Individuen in den Populationen (d.h. manche Populationen haben mehr Fliegen als andere)? Wenn nicht, woher stammen die Unterschiede in  $N_e$ ?

Der letzte Teil der heutigen Übung besteht darin mehr über die vier sequenzierten und analysierten Loci (55, 126, 530, 1435) aus *Drosophila melanogaster* herauszufinden. Benutzen Sie hierzu FlyBase (<http://www.flybase.org>).

4) Finden Sie zunächst heraus wo im Genom die vier Loci liegen. Benutzen Sie hierzu die BLAST-Suche von FlyBase und eine der Sequenzen die Sie für Ihre Loci generiert haben.

a) Auf welchem Chromosom liegt der Locus?

Locus	55	125	530	1435
Chromosom				

b) Wie lauten die genauen Anfangs- und Endkoordinaten der jeweiligen Loci? (Hinweis: Betrachten Sie die Koordinaten der „Subject“-Sequence Ihres BLAST-Treffers)

Locus	55	125	530	1435
Position				

5) Betrachten Sie nun das genomische Umfeld des Locus im Genom-Browser („GBrowse“). Ist er Teil eines proteinkodierenden Gens? Liegt er in einem Exon oder Intron? Um welches Gen handelt es sich und was für eine Funktion hat dieses Gen? Benutzen Sie hierzu die funktionellen Informationen zum Gen die in der FlyBase-Datenbank hinterlegt sind. (Hinweis: FlyBase-Übung von Woche 1).

Falls der Locus intergenisch ist, zwischen welchen proteinkodierenden Genen liegt er und welche Funktion haben diese Gene?

55	
126	
530	
1435	

6) Bei den Analysen haben Sie einen Locus identifiziert, der sich in einem genomischen Bereich befindet der unter Selektion stehen könnte (Frage 1). Dieser Locus wurde in *Drosophila* bereits umfassend dahingehend analysiert. Finden Sie heraus welches der Gene an diesem Locus im Verdacht steht das Ziel von natürlicher Selektion zu sein. Führen Sie dazu eine Google oder Literatursuche durch und suchen Sie nach einer Publikation.