

Evolutionsbiologie II für Bachelor und Lehramt - Übung 3, WS2016/2017

Anleitung für die Programme:

- GeneStudio Pro (<http://genestudio.com/>)
- DnaSP (<http://www.ub.edu/dnasp/> - Manual: <http://www.ub.edu/dnasp/DnaSPHelp.pdf>)

1) Assembly

Im ersten Schritt erstellen wir aus den Chromatogrammen die mit den Sequenzierreaktionen aus Woche 2 erzeugt wurden die Konsensussequenz aus der der jeweiligen Vorwärts- und Rückwärtsreaktion.

- Speichern Sie Ihre 8 Chromatogramme in einem Ordner auf dem Desktop. Die Dateien finden Sie im geteilten Ordner auf dem Rechner des Dozenten. Der Unterordner mit Ihrer Gruppennummer enthält Ihre Reaktionen.
- Sollten einige Reaktionen nicht funktioniert haben, stellen wir Ihnen Ersatzsequenzen zur Verfügung (markiert mit „Backup“).

Zusammenfügen der Chromatogramme:

- Öffnen Sie das Programm „GeneStudio Pro“ (→ GeneStudio,Inc → GeneStudio.exe).
- Klicken Sie auf „Contig editor“.
- Ziehen Sie die Vorwärts- und Rückwärtsreaktionen aus einem Individuum (z.B. 55F_E01.ab1 und 55R_E01.ab1) in den Contig editor („drag-and-drop“). Im Bereich „Viewers“ sollten beide Reaktionen sichtbar sein.
- Im erscheinenden Popup-Fenster wird gefragt, ob beide Sequenzen zusammengefügt werden sollen. Bestätigen Sie dies mit „Ja“.
- Bestätigen Sie im erscheinenden „Trim ends“ Fenster die Standardeinstellungen mit „OK“.
- Bestätigen Sie das nächste Fenster mit „OK“.
- Schließen Sie das „Contig control panel“ mit „Cancel“.
- Im Hauptfenster sollten jetzt beide zusammengeführten Reaktionen zu sehen sein. Vergewissern Sie sich zunächst ob Ihre Sequenzen die richtige Orientierung haben. Bei den farbigen Orientierungspfeilen muss der grüne Pfeil mit der Vorwärtsreaktion („F“) und der rote Pfeil mit der Rückwärtsreaktion („R“) verknüpft sein. Sollte dies nicht der Fall sein, kehren Sie die Reaktionen um indem Sie in Menü auf „Contig“ → „Reverse“ klicken.
- Machen Sie nun die Chromatogramme sichtbar, indem sie auf die schwarzen Pfeile links neben den Sequenznamen klicken.
- Mit den Reglern „Height“, „X scale“ und „Y scale“ können Sie die Größe der Chromatogramme ändern. Vergrößern Sie diese, damit Sie einen besseren Blick auf die Qualität der Sequenzen haben.

Editieren von Sequenzfehlern:

Vor allem am Anfang und am Ende von Sequenzierreaktionen sind die Signale nicht zuverlässig und es werden fehlerhafte Basen angezeigt. Diese werden meistens durch „Trim ends“ Funktion während des Assemblys erkannt und gekürzt, oft ist aber ein manuelles Nacharbeiten nötig. Um eine gesicherte Sequenz zu erhalten sollten alle Stellen von der Vorwärts- und der Rückwärtsreaktion abgedeckt sein und beide Sequenzreaktionen sollten überall die gleichen Basen anzeigen. Dies müssen wir nun überprüfen:

- Sorgen Sie zunächst dafür, dass beide Sequenzreaktionen den gleichen Bereich abdecken. Benutzen Sie dazu den kleinen schwarzen Pfeil am Ende der Sequenzreaktion um beide Reaktionen auf dieselbe Länge zu ziehen. D.h. verkürzen Sie den längeren Strang auf die Länge des kürzeren.
- Suchen Sie nun nach „Ambiguities“ (Uneindeutigkeiten) zwischen den beiden Strängen. Wenn beide Reaktionen unterschiedliche Basen an derselben Stelle anzeigen, so wird dies als rote Base in der Konsensussequenz über den Chromatogrammen sichtbar.
- Sie können im Menü mit „Contig“ → „Ambiguity finder“ gezielt nach solchen Stellen suchen.
- Solche „Ambiguities“ tauchen oftmals an den Enden der Sequenzen auf. In diesem Fall ist es am einfachsten beide Stränge so lange zu kürzen, bis diese „Ambiguities“ verschwinden. Dabei sollten jedoch nicht mehr als 20 bp gekürzt werden.
- Tauchen „Ambiguities“ in der Sequenzmitte auf, so kann dies oft mit einem erhöhten Hintergrundrauschen in einer der beiden Sequenzen erklärt werden. Betrachten Sie die Qualität der Chromatogramme an solchen Stellen. Oftmals ist eindeutig wo der Fehler liegt. Editieren Sie in diesem Fall den fehlerhaften Strang in dem Sie die Position markieren und die korrekte Base eintippen.
- Korrigieren Sie die Chromatogramme bis der „Ambiguity finder“ keine Fehler mehr anzeigt.

Speichern der Konsensussequenz:

- Geben Sie der Sequenz zunächst einen Namen. Gehen Sie im Menü auf „Contig“ → „Properties“ und ändern sie den Namen. Benutzen Sie dabei das Format *Locus_Individuum* (z.B. 55_E01).
- Speichern Sie die Konsensussequenz in dem Sie im Menü auf „Contig“ → „Export current consensus“ klicken
- Der Dateiname sollte dabei dem Namen Ihrer Sequenz entsprechen (z.B. 55_E01). Die Endung „.nt“ wird von Programm automatisch angefügt, die gespeicherte Datei wäre dann z.B. 55_E01.nt.
- Die Datei entspricht dem FASTA Format. Stellen Sie sicher, dass sie korrekt abgespeichert wurde indem Sie den Inhalt mit einem Text-Editor (z.B. Notepad++) betrachten.

Wiederholen Sie diesen Prozess für alle 4 Sequenzreaktionspaare. Nachdem Sie Ihre 4 Konsensussequenzen erhalten haben, kopieren Sie diese in den Sequenz-Ordner auf dem Rechner des Dozenten, damit die Sequenzen allen Gruppen im Kurs zugänglich sind.

2) Alignment

Um die DNA-Variabilität an den 4 sequenzierten Loci analysieren zu können, müssen zunächst Alignments mit allen Sequenzen erstellt werden. Dies geschieht mit der „Alignment editor“ Komponente von „GeneStudio Pro“.

- Vergewissern Sie sich zunächst das alle Sequenzen für einen Locus bereits im gemeinsamen Ordner des Dozenten vorhanden sind. Kopieren Sie dann den Ordner auf Ihren Desktop.
- Klicken Sie in GeneStudio Pro links auf die Komponente „Alignment editor“.
- Markieren und ziehen Sie alle Sequenzen eines Locus in das obere Fenster des „Alignment editor“ (drag-and-drop).
- Gehen Sie im Menü auf „Align Sequences“ → „Clustal W 1.83“ → „Fast/Approximate pair-wise alignment“.
- Ihr Alignment wurde nun erstellt. Überprüfen Sie visuell, ob das Alignment korrekt durchgeführt wurde. Die meisten homologen Stellen sollten monomorph sein und nur wenige Stellen sollten SNPs oder InDels aufweisen.
- Sollten sich eine oder mehrere Sequenzen extrem von den anderen unterscheiden, so kann dies daran liegen, dass Sequenzen aus falschen Loci benutzt wurden oder in der falschen Orientierung vorliegen. Verständigen Sie in diesem Fall einen Tutor, damit das Problem behoben werden kann.
- Sobald Sie das korrekte Alignment produziert haben, speichern Sie es ab. Gehen Sie dazu im Menü auf „File“ → „Save as...“.
- Setzen das „File format“ auf „FASTA“.
- Benennen Sie Ihre Datei nach dem Locusnamen und fügen Sie die Endung „.fas“ hinzu (z.B. 55.fas)
- Setzen Sie den „Export folder“ auf einen Ordner auf Ihrem Desktop und speichern Sie mit „OK“.
- Im erscheinenden Fenster „Gaps in sequence“ wählen Sie „Leave gaps“.
- Eventuell speichert das Programm Ihr Alignment dennoch mit der Endung „.nt“ (z.B. 55.nt). Benennen Sie in diesem Fall die Endung der Datei manuell auf dem Desktop in „.fas“ um.

Wiederholen Sie diesen Vorgang mit den Sequenzen aus allen 4 Ordnern und produzieren Sie somit Alignments für alle 4 Loci. Diese Dateien sind nun bereit für die populationsgenetische Analyse in DnaSP.

3) Populationsgenetische Analyse

- Starten Sie die Software DnaSPv5.
- Im Menü dieser Software gehen Sie nun auf „File“ → „Open Data File“ und öffnen das erste Alignment (z.B. 55.fas).
- Bevor Sie mit den Analysen beginnen können, müssen die Populationen definiert werden. Gehen Sie dazu im Menü auf „Data“ → „Define Sequence Sets“ und wählen Sie alle Individuen, welche zur afrikanischen *Drosophila melanogaster* Population gehören (A..). Sie können mehrere Individuen durch das gedrückt halten der „Shift“-Taste auswählen.
- Klicken Sie auf “>>” in der Mitte des Fensters - die ausgewählten Sequenzen sind nun im rechten Fenster (“Included List”).
- Klicken Sie auf „Add new sequence set“ und nennen Sie diese Auswahl „Afrika“. Ihre ausgewählten Sequenzen sind nun als Sequenzset „Afrika“ gruppiert.
- Wiederholen Sie diese Prozedur um alle Individuen aus den Populationen „Europa“ (E..) und „Asien“ (KL..) zu gruppieren. Abschließend erstellen sie eine Außengruppe „Out“, die nur die Sequenz von *D. simulans* (Sim1) enthält.
- Im Auswahlmenü „Sequence Sets“ unten links sollten nun alle 4 Gruppen auftauchen. Klicken Sie auf „Update All Entries“ um den Prozess abzuschließen. Der Datensatz ist nun bereit für die Analysen.
- Sie könne Ihren Datensatz betrachten in dem Sie im Menü auf „Display“ → „View Data“ klicken.
- Wählen Sie im erscheinenden Alignment-Fenster unter „Select Sites/Codons...“ nun „Variable (polymorphic) sites“. Dadurch werden alle Stellen mit SNPs farblich markiert. Der Großteil Ihres Alignments sollte jedoch monomorph sein.

Um die Fragen auf Ihrem Arbeitsblatt zu beantworten müssen Sie zunächst mehrere populationsgenetische Statistiken berechnen. Es handelt sich hierbei um:

| | |
|------------------------|---|
| n: | Stichprobengröße |
| bp: | Länge des verwendeten Alignments |
| S: | Anzahl an polymorphen Stellen |
| θ_w (per site): | Waterson's Schätzer |
| π (per site): | Durchschnittliche Anzahl an Unterschieden |
| Tajima's D: | Tajima's Statistik |
| K: | Divergenz zwischen Arten |
| F_{ST} : | Populationsdifferenzierung |

Um diese Statistiken zu erhalten werden mehrere Module in DnaSP benutzt:

- Gehen Sie im Menü auf „Analysis“ → „Tajima's Test...“.
- Wählen Sie im erscheinenden Fenster bei „Data Set:“ eine Ihrer definierten Populationen (z.B. Afrika). Somit werden die Statistiken für Ihre gewählte Population berechnet.
- Wählen Sie unter „Nucleotide Substitutions Considered“ den Punkt „Segregating sites“ und klicken Sie auf OK.

Das Ausgabefenster enthält nun die gewünschten Statistiken:

n: Number of sequences used
bp: Total number of sites (excluding sites with gaps / missing data)
S: Number of polymorphic (segregating) sites, S
 θ_w (per site): Theta (per site) from S, Theta-W
 π (per site): Nucleotide diversity, Pi
Tajima's D: Tajima's D

Tragen Sie die Werte in die Tabelle „Summary statistics“ in Ihrem Arbeitsblatt ein und wiederholen Sie den Vorgang für die zwei verbliebenen Populationen. Um die Divergenz zu erhalten benötigen wir ein anderes Modul:

- Gehen Sie im Menü auf „Analysis“ → „Polymorphism and Divergence...“ und bestätigen Sie mit „OK“.
- Wählen Sie im erscheinenden Fenster unter „Data Sets“ für „Intraspecific Data“ eine Ihrer Populationen (z.B. Afrika) und für „Interspecific Data“ die Außengruppe „Out“. Klicken Sie auf OK.
- Die Divergenz (K) befindet sich im Ausgabefenster unter „Nucleotide Divergence, Ks (Silent)“

Wiederholen Sie den Vorgang für die beiden verbliebenen Populationen und tragen Sie die Werte in die Tabelle ein. Abschließend berechnen wir F_{ST} -Werte:

- Gehen Sie im Menü auf „Analysis“ → „Gene Flow and Genetic Differentiation...“.
- Im erscheinenden Fenster sollte unter „Net Number of Populations Included“ die Zahl 3 stehen. D.h. das Programm wird alle drei Populationen miteinander vergleichen. Klicken Sie auf „OK“.
- Im Ausgabefenster „Genetic Differentiation Among Populations“ finden Sie Statistiken für die drei paarweisen Vergleiche. Finden Sie die Spalte mit den F_{ST} -Werten und tragen Sie diese in die Tabelle ein.

Nachdem Sie alle Statistiken berechnet haben, schließen Sie Ihr Alignment, indem Sie im Menü auf „File“ → „Close Data File“ klicken. Wiederholen Sie die Berechnung der Statistiken für die anderen Loci und füllen Sie die Tabellen im Arbeitsblatt aus.

Beantworten Sie die Fragen auf dem Arbeitsblatt und löschen Sie abschließend die von Ihnen produzierten Dateien von Ihrem Rechner.