

## Ćwiczenia nr 4.

### *Programy komputerowe stosowane do analizy danych molekularnych*

Istnieje ogromna liczba programów służących do analizy danych molekularnych. Wiele z nich zostało napisanych tylko dla jednej, konkretnej metody i dlatego znajdują zastosowanie jedynie w wyspecjalizowanych analizach. Jednak nawet liczba programów oferujących wiele metod jest znaczna i cały czas pojawiają się nowe, co stanowi źródło zamieszania i wielu frustracji. Poniżej przedstawiono dwa programy, które ze względu na łatwość użycia i szeroki wachlarz oferowanych metod uznaliśmy za najbardziej przydatne i najbardziej reprezentatywne (stan: listopad 2011 r.).

#### **I. Arlequin – analiza allelicznych danych populacyjnych**

Arlequin pozwala na analizy genetyczno-populacyjne wszelkich typów danych molekularnych: allozymów, mikrosatelitów, polimorfizmów pojedynczych nukleotydów (SNP), markerów dominujących RAPD i AFLP, wzorców cięcia enzymami restrykcyjnymi (RFLP) oraz sekwencji DNA.

W przygotowanym pliku znajdują się dane o genotypach 4 genów MHC II w 7 populacjach traszki górskiej z Polski.

Za pomocą programu Arlequin obliczymy wartości  $F_{ST}$  dla każdej pary populacji i sprawdzimy ich istotność statystyczną.  $F_{ST}$  to miara zróżnicowania genetycznego populacji, wyrażana jako wariancja częstości alleli między populacjami; przyjmuje wartości od 0 do 1, 0 oznacza zupełny brak zróżnicowania między populacjami. Hipotezę zerową iż  $F_{ST} = 0$  Arlequin testuje testami randomizacyjnymi.

Prawo Hardy’ego-Weinberga mówi iż w nieskończenie dużej populacji organizmów rozmnażających się płciowo, gdy kojarzenie jest losowe, pokolenia nie zachodzą na siebie, brak jest mutacji, migracji oraz doboru naturalnego lub też te trzy czynniki równoważą się, częstości alleli i genotypów nie będą zmieniały się z pokolenia na pokolenie, a częstości genotypów można obliczyć znając częstości alleli.

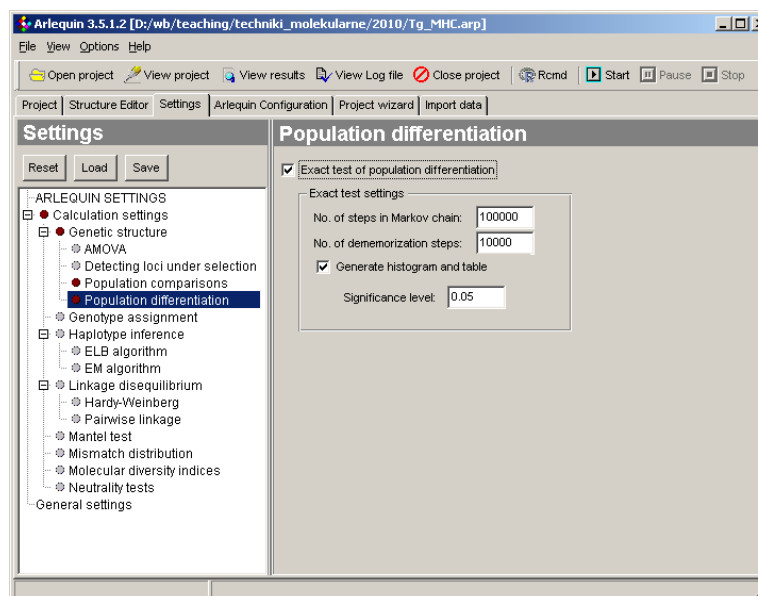
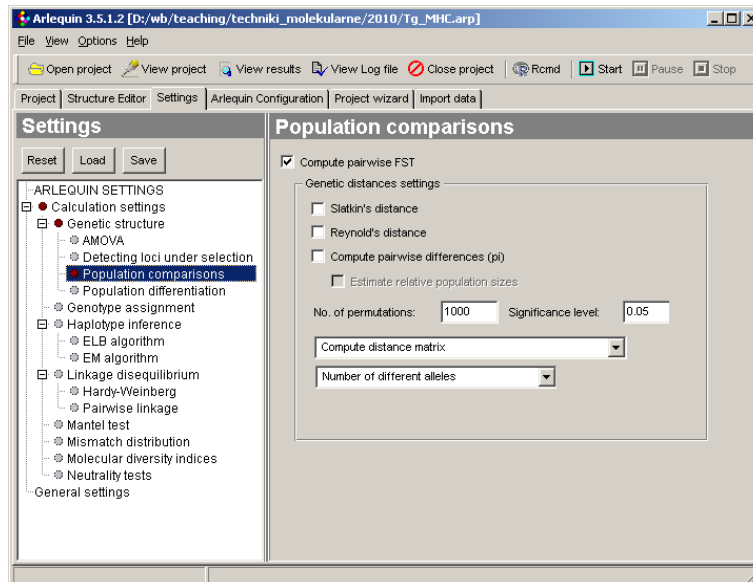
W populacjach naturalnych zazwyczaj obserwuje się dobrą zgodność obserwowanej częstości genotypów z oczekiwanymi z prawa Hardy’ego-Weinberga. Dzieje się tak dlatego, iż jedno pokolenie losowego kojarzenia przywraca zgodność częstości obserwowanych i oczekiwanych, podczas gdy czynniki wywołujące odchylenia, jak dryf genetyczny czy dobór naturalny działają zazwyczaj w dłuższej skali czasowej. A zatem może być tak, że częstości alleli zmieniają się z pokolenia na pokolenie lecz jeżeli tylko kojarzenia są losowe, w każdym pokoleniu będziemy obserwować zgodność z prawem Hardy’ego-Weinberga. Hipotezę o braku odchylenia od równowagi Hardy’ego-Weinberga testujemy symulacjami Monte Carlo. Program generuje próby o liczebności i częstościach alleli takich jak w naszych danych, lecz przy założeniu równowagi Hardy’ego-Weinberga i sprawdza w jakiej części z tych prób odchylenia od wartości oczekiwanych będą większe niż dla naszych danych. Na podstawie tego wyniku oblicza istotność statystyczną testu.

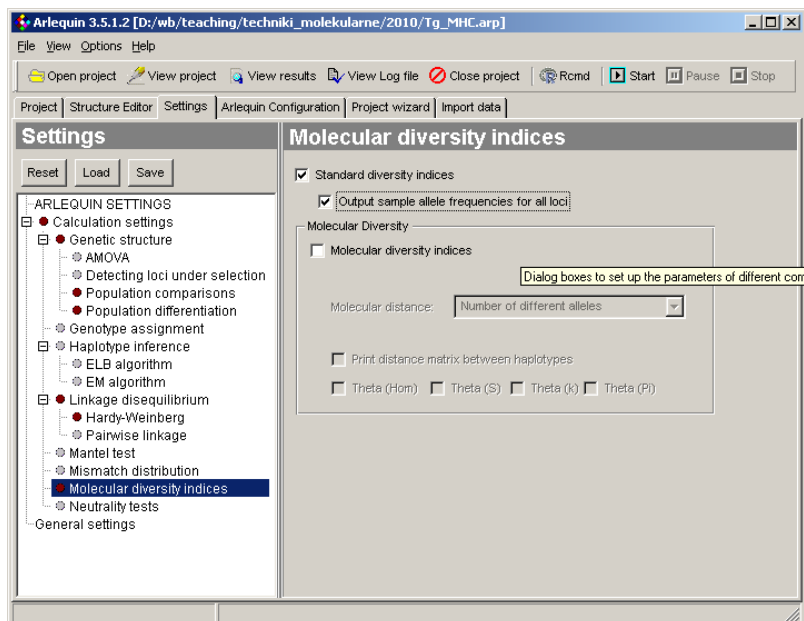
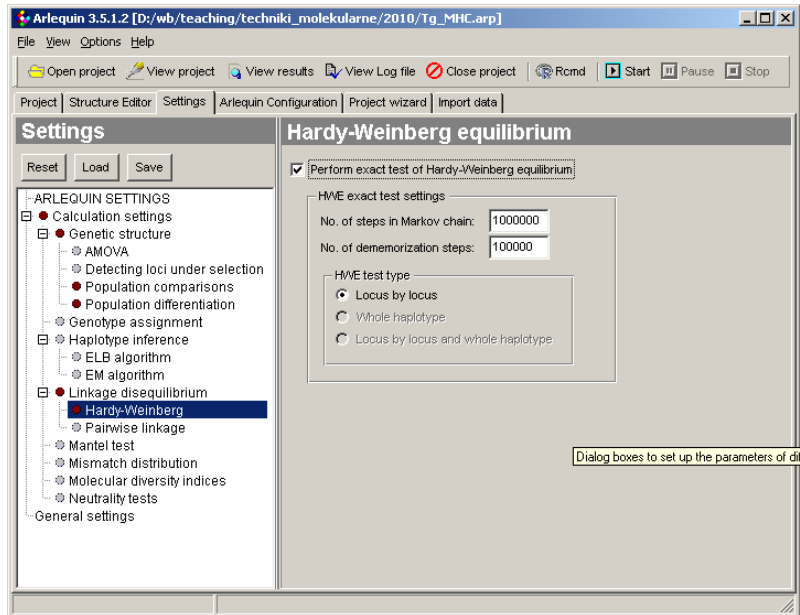
Omówione wyżej procedury to standardowe pierwsze kroki w analizie danych allelicznych.

1. Uruchom program WinArl35 i w karcie ‘Arlequin configuration’ ODZNACZ ‘XML Output’.

2. Wybierz Open Project i otwórz projekt 'Tg\_MHC.arp'

3. Przejdź do zakładki 'Settings' i ustaw odpowiednie parametry jak poniżej:





4. Naciśnij 'Run', postęp obliczeń widać w lewym dolnym brzegu okienka.
5. Po ukończeniu obliczeń wyniki zostaną zapisane w folderze Tg\_MHC.res w kilku plikach, w przeglądarce otwieramy Tg\_MHC\_main.htm.
6. Odpowiedz na pytania:

- między którymi populacjami  $F_{ST}$  jest **największe**?
- między którymi populacjami  $F_{ST}$  **nie jest istotne statystycznie**, co to oznacza ze **statystycznego** a co z **biologicznego** punktu widzenia?
- między którymi populacjami **zróźnicowanie częstości genotypów** jest statystycznie istotne?
- w części '*Hardy-Weinberg equilibrium*' dla każdej populacji porównaj obserwowane (*Obs. Het.*) oczekiwane (*Exp. Het.*) heterozygotyczności oraz wyniki testów na

równowagę Hardy'ego-Weinberga (*P-value*) dla poszczególnych loci w poszczególnych populacjach? Czy wszystkie loci we wszystkich populacjach są w równowadze Hardy'ego-Weinberga? A jeżeli nie to gdzie mamy odchylenia? O czym mogą one świadczyć



- w części '*Summary of computations done within populations*' porównaj liczbę alleli w poszczególnych loci, o czym mogą świadczyć obserwowane drastyczne różnice?


## II. MEGA5 – analiza sekwencji DNA

Program MEGA5 używany jest do konstrukcji i rysowania drzew filogenetycznych, szacowania dywergencji sekwencji, testowania działania doboru naturalnego na sekwencje kwasów nukleinowych i białek. Jest nowoczesnym i elastycznym, a przy tym bardzo prostym w użyciu narzędziem.


W ćwiczeniu będziemy analizować 10 sekwencji fragmentu genu NCX (koduje wymiennik jonowy wapń-sód) u trzaski zwyczajnej, karpackiej i grzebieniastej. Obejrzymy sekwencje nukleotydów i aminokwasów oraz skonstruujemy drzewo filogenetyczne obrazujące ich podobieństwa/pokrewieństwa.

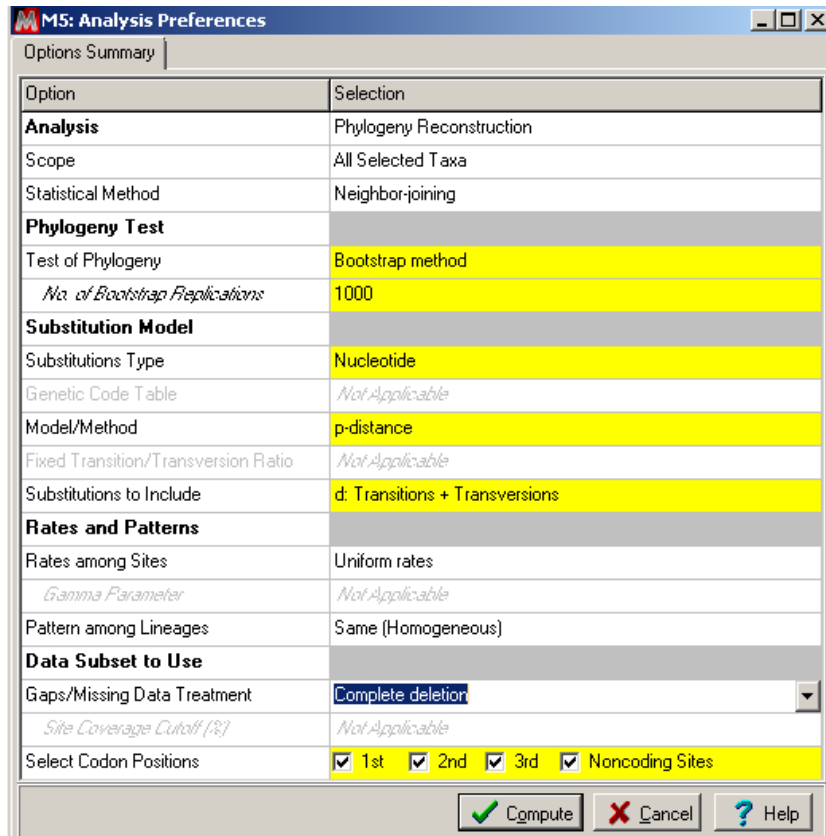
1. Uruchom program Mega4
2. Wybierz *File* -> *Open A File/Session...*
3. Otwórz plik NCX.meg

4. Kliknij przycisk 
- 5., Plik pojawi się w oknie przeglądarki, kropki oznaczają że dany nukleotyd jest identyczny we wszystkich sekwencjach. Aby zobaczyć całą długość sekwencji przewijaj w oknie
6. Kliknij przycisk , przełącza on między widokiem z kropkami a widokiem wszystkich sekwencji

7. Klikaj teraz przyciski z tej grupy 
  - C – podświetla miejsca niezmiennie
  - V – podświetla miejsca zmienne
  - S – podświetla mutacje występujące jedynie raz w zbiorze danych
  - 0 – podświetla miejsca niezdegenerowane, czyli takie w których dowolna zmiana spowoduje zmianę aminokwasu
  - 2 – podświetla miejsca dwukrotnie zdegenerowane, czyli takie w których tylko niektóre zmiany spowodują zmianę aminokwasu
  - 4 – podświetla miejsca czterokrotnie zdegenerowane, czyli takie w których żadna zmiana nie spowoduje zmiany aminokwasu

W której z kategorii jest najmniej miejsc zmiennych? O czym to świadczy?

8. Kliknij przycisk , spowoduje to wyświetlenie sekwencji aminokwasów zamiast nukleotydów, czy zmienność aminokwasów jest taka sama, większa czy mniejsza niż zmienność nukleotydów? O czym to świadczy?
9. Zamknij okno 'Data Explorer'
10. Wybierz z głównego Menu: *Phylogeny* -> *Construct/Test Neighbor Joining Tree...* i ustaw parametry następująco:



11. Kliknij 'Compute', po krótkiej chwili w nowym oknie pokaże się skonstruowane drzewo. Co mówią nam długości gałęzi a co liczby nad niektórymi gałęziami?
12. Przyciskami po lewej stronie i u góry ekranu możesz manipulować wyglądem drzewa, zwróć uwagę że **cały czas jest to samo drzewo**, zmienia się jedynie sposób jego graficznej prezentacji.
13. Skonstruuj drzewo Neighbor-Joining używając procentu różnic aminokwasowych jako miary odległości genetycznej. Czy otrzymane drzewo różni się od tego uzyskanego z sekwencji nukleotydów? Dlaczego?