

Differentiering i en linjär population, en matematisk undersökning

Mattias Siljestam

Inom populationsgenitiken har utvecklats modeller för att mäta differentiering mellan populationer varit ett aktivt forskningsområde sedan Sewall Wright publicerade fixeringsindexet F_{ST} år 1943. F_{ST} är ett statistiska som har ett värde mellan noll och ett och beskriver graden av differentiering mellan populationer. Detta statistiska används för att undersöka arters historia genom att mäta hur differentierade dess populationer är idag. Mätning av differentiering mellan population kan användas för att hjälpa oss förstå hur forntida migrationer skedde, exempelvis hur *Homo sapiens* spreds sig ut från Afrika en gång i tiden. Det kan även användas för undersöka i vilka gener naturlig selektion sker och har skett.

Men F_{ST} har inte blivit undersökt för scenariot av en linjär population, det vill säga en population som består av ett flertal subpopulationer som ligger längs en linje. En linjär population kan liknas vid en kedja, där varje länk utgör en subpopulation. En modell för en sådan linjär population skulle i många sammanhang kunna matcha verkligheten bättre och därmed vara mer lämplig att använda än de modeller som existerar idag.

Denna uppsats fokuserade på att utveckla en verktygslåda som undersöker differentiering för linjära populationer. Två andra statistiska inkluderades utöver F_{ST} : Jots D_{ST} och Hedricks standardiserade G_{ST} . Dessa inkluderades för att beskriva differentiering från flera vinklar, vilket ger en mer komplett bild.

Dessa tre statistiska beräknades för linjära populationer av olika storlekar, men beräkningarna gav en uppsjö av gigantiska ekvationer som svar. Men jag lyckades från alla dessa stora och svårtolkade ekvationer få fram tre små och häandiga approximerade ekvationer, en för varje statistiska. Dessa tre approximerade ekvationer var enkla att förstå och använda och de beskrev mönstret bakom de jättelika ekvationerna med en näst intill perfekt matchning.

De tre approximerade ekvationerna utgjorde min verktygslåda och kunde beskriva relationen mellan den förväntade differentieringen och de tre viktigaste mekanismerna i populationsgenetik: mutations- och migrationsfrekvens samt populationsstorlek. Med hjälp av dessa tre ekvationer kunde även den absoluta migrationsfrekvensen estimeras, vilket få modeller idag har möjligheten att göra.