

УДК 575 : 004.94

М.Прокоп'як, канд. біол. наук, доц.

Г.Гуменюк, канд. біол. наук, доц.

Н.Дробик, докт. біол. наук, проф.

Тернопільський національний педагогічний університет імені Володимира Гнатюка,
Україна

АНАЛІЗ ГЕНЕТИЧНОЇ СТРУКТУРИ ПОПУЛЯЦІЙ РОСЛИН З ВИКОРИСТАННЯМ ПРОГРАМ, ЯКІ БАЗУЮТЬСЯ НА ЙМОВІРНІСНІЙ МОДЕЛІ БАЙЕСА

M. Z. Prokopiak, Ph.D., Assoc. Prof.,

H. V. Humeniuk, Ph.D., Assoc. Prof., N. M.

Drobyk, D.Sc., Prof.

Volodymyr Hnatyuk Ternopil National Pedagogical University, Ukraine

ANALYSIS OF GENETIC STRUCTURE OF PLANT POPULATIONS USING BAYESIAN PROBABILITY MODELS

Існують різні програми для моделювання генетичних даних і генетичних процесів у популяціях. Structure – статистична програма кластерного аналізу для вивчення структури популяцій, яка використовує полілокусні генотипні дані. Програма створена на основі математичного алгоритму, описаного (Pritchard et al. 2000), в якому використовується Байєсовський алгоритм; ця програма ґрунтується на кількісному кластерному аналізі. За допомогою Structure можна виділяти чіткі популяції, оцінювати належність індивідів до певної популяції. Аналіз у Structure може здійснюватися для різних молекулярних маркерів (включаючи SNP, SSR, RFLP, AFLP й ін.), а також можна використовувати фенотипові дані.

Ще одне програмне забезпечення Geneland призначене для аналізу генетичних даних та виявлення популяційної структури на основі молекулярних маркерів. Ця програма також ґрунтується на алгоритмі Монте-Карло за схемою марківських ланцюгів для байєсівської статистики. Geneland працює з даними, отриманими за кодомінантними і домінантними маркерами, а у версії 4.0.0 можна завантажувати фенотипові дані. Програма допомагає визначити, наскільки генетичні дані вказують на наявність різних популяцій або генетичних груп у досліджуваній вибірці.

У програмі Structure для аналізу даних популяційної генетики використовуються методи Марковських ланцюгів Монте-Карло (МСМС). МСМС є потужним інструментом для вирішення складних статистичних задач і використовується в багатьох галузях, включаючи генетику та біоінформатику, для оцінки параметрів і проведення аналізу в умовах невизначеності. Марковські ланцюги Монте-Карло використовуються в Structure для оцінки ймовірностей різних генетичних структур популяції і дають можливість зробити висновки про розподіл генів серед різних підгруп популяції. Цей процес повторюється багато разів (велика кількість ітерацій), і на основі отриманих результатів можна робити статистичні висновки щодо популяційної структури.

Нами для побудови генетичної структури шести популяцій рослин тирличу жовтого (*Gentiana lutea* L.) було використано програму Structure 2.3.4. Вона адаптована для домінантних маркерів, тому нами проведено молекулярно-генетичний аналіз методами RAPD-, ISSR-, IRAP-, RGAP- і CDDP-ПЛР. Вони дозволили оцінити велику кількість локусів одночасно і, відповідно, за ними побудувати генетичну структуру популяцій. Для аналізу було використано модель, яка допускає змішане походження особин, передбачає кореляцію частот алелей у батьківських популяціях (K) і не враховує приналежності особин до географічних локалітетів популяцій. Для домінантних маркерів «відсутні» значення даних (missing data value) задавали як -9. Для визначення найбільш вірогідного числа K виконано серію аналізів від 1 до 9 (20 повторностей на K), з періодом припрацювання (a burnin period)

50000 і 300000 ітерацій (а Markov chain Monte Carlo (MCMC)) (Pritchard et al., 2000; Falush et al., 2007). Для визначення оптимального значення К використано метод Evanno (2005).

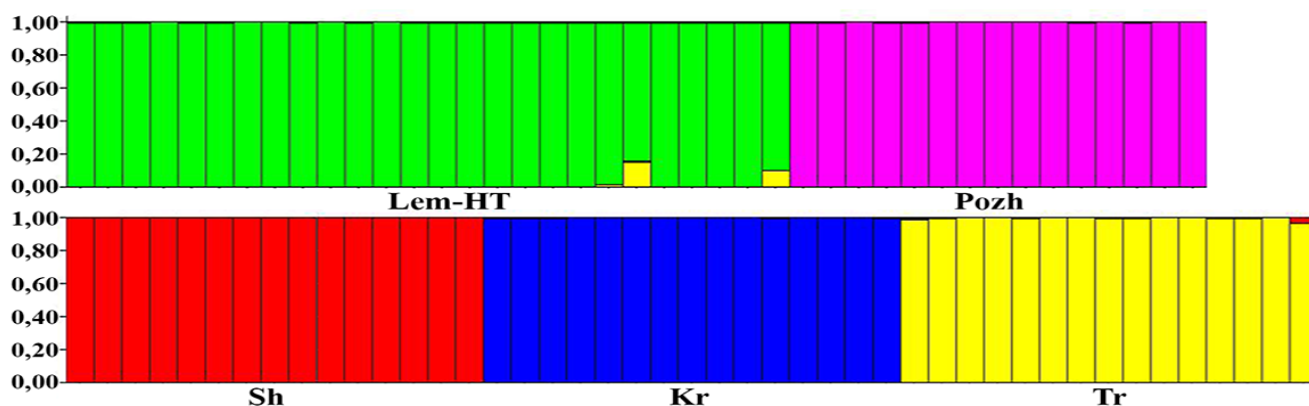


Рис.1. Структура генетичної різноманітності *G. lutea* з Українських Карпат (гг. Шешул-Павлик (Sh), пол. Лемська (Lem), г. Гутин Томнатик (NT), гг. Трояска-Татарука (Tr), пол. Крачунеска (Kr)), г. Пожижевська (Pozh)) на основі RAPD-, ISSR-, IRAP-, RGAP- і CDDP-маркерів з використанням програми Structure 2.3.4.

Згідно даних цього аналізу, лемська і гутин томнатиська популяції мають спільне походження. В об'єднаній популяції «Lem-NT» виявлено три особини, які містять генетичний матеріал із трояської-татарука популяції (1,7 %, 10,4 % та 15,6 %). У трояської-татарука популяції виявлено особину із спільним генетичним матеріалом від шешульської у кількості 3,2 %. В усіх кластерах виявлено особини спільного походження у залишкових кількостях (0,1–0,5 %).

Сьогодні існує чимало аналогічних програм. Наприклад, програма BEAST (Bayesian Evolutionary Analysis Sampling Trees) – програмне забезпечення для філогенетичного аналізу на основі байєсівської статистики. Воно використовується для оцінки еволюційних дерев, часу розділення гілок й інших параметрів еволюційних процесів на основі генетичних даних. BEAST є потужним інструментом для реконструкції філогенетичних дерев і оцінки еволюційних параметрів із використанням байєсівських методів. BEAST зазвичай використовується для аналізу генетичних даних на основі ко-домінантних маркерів.

Отже, для побудови генетичної структури популяцій і аналізу генетичних даних часто використовують математичні методи. Чимало математичних підходів можуть бути використані для аналізу генетичних даних, визначення генетичної структури популяцій і розуміння еволюційних процесів. Однак важливо враховувати походження даних, мету і поставлені завдання для правильного вибору відповідних математичних методів і моделей.

Література

1. Evanno G., Regnaut S., Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE : a simulation study. *Molecular Ecology*. 2005. Vol. 14, Is. 8. P. 2611–2620.
2. Falush D., Stephens M., Pritchard J. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes*. 2007. Vol. 7, № 4. P. 574–578.
3. Pritchard J., Stephens M., Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics*. 2000. Vol. 155. P. 945–959.