*Genetics of Aquatic Organisms 1: 71-79 (2017)*
www.genaqua.org
ISSN 2459-1831
DOI: 10.4194/2459-1831-v1_2_04

REVIEW

# Analysis of Microsatellite Data for Population Genetics of Aquatic Organisms

**Rafet Cagri Ozturk[1], Ilhan Altinok[1,*]**

[1] Karadeniz Technical University, Surmene Faculty of Marine Sciences, Department of Fisheries Technology Engineering, 61530 Surmene, Trabzon, Turkey.

* Corresponding Author: Tel.: +90.462 3778030;
E-mail: ialtinok@ktu.edu.tr

Received 22 November 2017
Accepted 25 December 2017

**Abstract**

Genetic analysis is vital for conservation biology, stock management and for understanding evolutionary processes at a population level. Microsatellites are one of the most popular genetic markers in molecular ecology and population genetics, conservation and wildlife management. Microsatellites or simple sequence repeats are distributed throughout the genome of eukaryotes and prokaryotes and found in both coding and non-coding nuclear and organellar DNA. Non-coding microsatellites are ideal genetic markers to estimate population structure, gene flow, genetic diversity, genetic distance in wild population and to assess level of inbreeding, parentage analysis and to map quantitative traits in the domesticated and cultivated species. To study microsatellite loci, following workflow has to be completed. Isolation of DNA from small non-lethal tissue (fin clip) sample, amplification of microsatellite regions with multiplex PCR using fluorescent dye labelled primers, making of electropherogram of fragments with analyzer and analysis of the data with the population genetic software packages which enable to check data errors, calculate basic population parameters and sophisticated analyses. In this review, we described features and functionalities of programs useful for microsatellite data analysis.

*Keywords*: Population genetic, genetic programs, error-checking, genetic parameters.

## Introduction

Application of genetic techniques such as microsatellites has blazed a trail in the fields of evolution, ecology, conservation and wildlife management. Microsatellites, along with single nucleotide polymorphism (SNP) are among the most popular genetic markers in molecular ecology (Ellis *et al*., 2011). Microsatellites or simple sequence repeats (SSRs) are arrays of short tandemly repeated motifs distributed throughout the genome of eukaryotes and prokaryotes and found in both coding and non-coding nuclear and organellar DNA (Phumichai, Phumichai, & Wongkaew, 2015). Due to their elevated mutation rate ranging from $10^3$ to $10^6$ per cell generation, these markers are highly variable (Gemayel, Cho, Boeynaems, & Verstrepen, 2012). The occurrence of microsatellites in gene coding regions is lower because of the high mutation rate which could compromise gene expression (Hoffman & Amos, 2005). The high polymorphism of these genetic markers, combined with their putative selective neutrality, Mendelian inheritance and straightforward assay compatible with high throughput genotyping of samples with the possibility of multiplexing loci (Renshaw, Saillant, Bradfield, & Gold, 2006), has led high popularity for genetic studies of various organisms (Sunnucks, 2000). Microsatellite markers are widely used for wild species to assess; (i) effective population size (Hauser, Adcock, Smith, Ramirez, & Carvalho, 2002), (ii) genetic diversity and genetic distance (Baillie, Muir, Scribner, Bentzen, & Krueger, 2016), (iii) gene flow (McKeown, Hauser, & Shaw, 2017) and (iv) population structure (de Vasconcellos, Lima, Bonhomme, Vianna, & Sole-Cava, 2015). On the other hand, for the domesticated and cultivated species, microsatellites are useful for (i) parentage analysis (Harrison *et al*., 2014), (ii) assessing level of inbreeding (Hillen *et al*., 2017), (iii) mapping quantitative traits (Lv *et al*., 2016) and (iv) constructing linkage maps (Hollenbeck, Portnoy, & Gold, 2015). To study microsatellite loci, DNA from a tissue sample has to be extracted, amplified, and DNA fragments have to be separated by capillary electrophoresis and microsatellites analyzed with suitable software.

### Work Flow of Microsatellites Scoring

Genomic DNA isolation is a fundamental step to study population genetics. DNA isolation method needs to be efficient, and economic. Application of

non-destructive and non-lethal protocols should be preferred especially for the endangered species. Fish fins for instance are desirable source of DNA and proved to be non-destructive. Standard PCI method (Phenol, Chloroform, Isoamyl alcohol) method and commercial DNA extraction kits (Qiagen, Promega) are suitable for isolating DNA from fin clips and 50-100 mg of tissue is enough in most of the cases. Once genomic DNA is extracted, quality and integrity of DNA should be verified by electrophoresis in a 1% agarose gel prior to PCR. Success of PCR amplification of microsatellites depends on template DNA quality. Degradation, low purity or quantity of DNA may result failure in amplification of longer alleles and may also cause artifact bands.

Generally large number of microsatellites are employed for population genetic studies. Performing single PCR reaction for each sample with each marker often increases cost and time consumption. Multiplex polymerase chain reaction (mPCR), simultaneous co-amplification of two or more loci in a single PCR reaction, is a technique that reduces the time and cost considerably in genetic analysis. Even eight to ten loci can be amplified in a same mPCR reaction. Each primer (either reverse or forward) labeled with fluorescent dyes (Hex, Fam, Ned, etc.) simultaneously amplify desired microsatellite loci and allow detection of targeted alleles in different loci in the same electropherograms. While, microsatellites with non-overlapping allele size ranges can be labeled with the same fluorescent dye, microsatellites with overlapping allele-size need to be labeled with different fluorescent dyes. Before labeling primers with fluorescent dyes, annealing temperatures need to be optimized and approximate range in allele sizes for each microsatellite needs to be known. Primers with similar optimum annealing temperatures need to be combined in each mPCR assay. Multiplex manager (Holleley & Geerts 2009) is a suitable tool that designs and optimizes mPCRs. This tool relies on the information given by user; sequence of primers, dye preferences, allele size range, annealing temperature and generates possible mPCR sets, which gives an idea and also a point to start.

The assay of microsatellites is based on the sizing of PCR fragments on acrylamide gels, which allows inferring the number of repeats of the two alleles in an individual. This assay can be easily developed when sequences of PCR primers specific to the flanking region of the microsatellite are available. The discovery of microsatellite loci and design of specific PCR primers has however been traditionally challenging. Methods to discover microsatellite alignments traditionally involved generating genomic libraries (typically following enrichment of genomic fragments containing specific microsatellite loci), cloning the obtained genomic fragments and screening the library for microsatellite arrays. This tedious process can now be avoided by next-generation sequencing methods, in particular, the direct Seq-to-SSR methods developed by (Castoe *et al*., 2012). This method exploits paired-end genome sequencing of a genomic library and it can be developed rapidly and cost-effectively for a variety of organisms (Antoni, Luque, Naghshpour, & Saillant, 2014; Li, Zhao, Peng, & Zhang, 2016). The software PAL-finder developed also by Castoe *et al*. (2012) allows screening of raw paired-end reads, without the need to assemble reads in contigs and design primers from the flanking regions.

Numerous population genetic software programs and tools are available online to analyze microsatellite data and most of them are free of charge. While some of the programs are designed for specific statistical analysis, others are programmed to calculated basic parameters. Same or similar parameters can be generated by different programs. The choice of a program will depend on individual preferences, availability, accessibility and interface. In this review, we described some of the population genetic software programs and tools, which are commonly employed in population genetic analysis. The programs included in this review and online links of programs are listed in Table 1. We also assess issues, common source of errors and consequences of errors in this review.

**Genotyping and Loci Scoring**

Microsatellite genotyping is a DNA characterization technique and an initial step of the data analysis. There are several programs for fragment analysis of microsatellite electropherograms. While some of them offer demo version for free with restricted use, most of them such as GeneMarker (SoftGenetics, LLC), GeneMapper (Applied Biosystems) are commercial and researchers have to purchase them for unrestricted use. There are also free software programs such as STRand (University of California Davis) which can be used for peak size identification.

Once sample files are received from fragment analyzer, screening of a locus across multiple samples would reveal alleles in expected size range and most importantly it would give an idea about peak pattern characteristics. Understanding the peak pattern characteristic of a locus would ease the process of allele scoring. Very often shape of peaks will be different between loci. For example, while a locus exhibits very clear single peak, another one may exhibit more than one major peak per allele. Some of the software programs have automated allele detection and calling feature, which gives a general idea, but the output is not 100% trustable. Un-optimized PCR, pipetting error, contamination, low quality DNA template, variation in thermal cycling conditions and other factors may cause variation and would possibly cause differences in peak height. Automatic allele calling with peak height restriction may cause failure and researcher may overlook some of the true alleles. Verifying each allele by hand is the

**Table 1.** List of the reviewed population genetic software programs

| Software | Input Data Format | Description of Functionality | Link | References |
|---|---|---|---|---|
| Convert | *.tab delimited text, Genepop | File Conversion | https://www.agriculture.purdue.edu/fnr/html/faculty/Rhodes/Students%20and%20Staff/glaubitz/software.htm | (Glaubitz, 2004) |
| Formatomatic | *.pnr, csv write, Genepop | File Conversion | http://taylor0.biology.ucla.edu/~manoukis/Pub_programs/Formatomatic/ | (Manoukis, 2007) |
| MICROCHECKER | *.txt, *.xlsx, Genepop | Detection and identification of genotyping errors; null alleles, allele dropout and allele stuttering | http://micro-checker.software.informer.com/2.2/ | (Van Oosterhout *et al*., 2004) |
| GENEPOP | Genepop | Computes genetic diversity indices, F-statistics, exact test of LD and HWE | http://kimura.univ-montp2.fr/~rousset/Genepop.htm. | (Raymond & Rousset, 1995) |
| Fstat | *.dat | Computes genetic diversity indices, allelic richness, F-statistics, mantel test, HWE test | https://www2.unil.ch/popgen/softwares/fstat.htm | (Goudet, 1995) |
| GenAlEx | genotype file in excell | Computes genetic diversity indices, Mantle test, PCoA, Fstatitstics, Nei's genetic distance measures. | http://biology-assets.anu.edu.au/GenAlEx/Welcome.html | (Peakall & Smouse, 2006) |
| Arlequin | *.arp | Computes genetic diversity indices, F-statistics, mantel test, exact test of LD and HWE, hierarchical analysis of genetic structure | http://cmpg.unibe.ch/software/arlequin35/Arl35Methods.html | (Excoffier & Lischer, 2010) |
| CERVUS | Specific | Parantage and identity analysis | http://www.fieldgenetics.com/pages/login.jsp | (Kalinowski *et al*., 2007) |
| STRUCTURE | *.str | Detect recent immigrants and genetic structure | https://web.stanford.edu/group/pritchardlab/structure.html. | (Pritchard *et al*., 2000) |
| TESS | *.str | Detect recent immigrants and genetic structure | http://membres-timc.imag.fr/Olivier.Francois/tess.html. | (Chen *et al*., 2007) |
| SAMOVA | *.arp and *.geo | Defines groups of populations | http://cmpg.unibe.ch/software/samova2/ | (Dupanloup *et al*., 2002) |
| Bottleneck | Genepop, *.dat | Identifies populations that experienced secere reduction in population size | http://www1.montpellier.inra.fr/CBGP/software/Bottleneck/bottleneck.html | (Piry *et al*., 1999) |
| LDNe | Genepop | Estimates effective population size | http://conserver.iugo-cafe.org/user/Robin%20Waples/LDNe | (Waples & Do, 2008) |

best way to score alleles. Besides that, most of the errors are derived from PCR conditions, specificity, and features of the primers. PCR optimization is the key in success of allele scoring; even in some cases, re-extraction of template DNA increases the accuracy of scoring. Another issue is that if too much PCR product is run, it may cause high fluorescence signal, which may overlap with any primer dye or may result in artifact peak known as "pull-up". (Figure 1) Mutation at primer binding site, alleles with a short length and PCR failure due to low quality or quantity of DNA may cause null alleles or allelic dropout (Figure 2). Inconsistent genotypes need to be re-analyzed before making the last decision. It is worth to note that allele scoring is an important task. The way the population genetic analysis going would change either in a good way or bad way based on scoring.

## Software Used for Microsatellite Analysis

### Data Conversion

Usage of a conversion program is an initial step of analysis after scoring alleles and obtaining raw microsatellite data set. Analysis of a microsatellite data typically requires use of a vast variety of software program and each with a different input data format. Data format conversion was once a time consuming and error-prone process. Thanks to conversion software programs, even large data sets can be converted into required input format within seconds. There are different types of free file converter software such as CONVERT, FORMATOMATIC, CREATE, and PGDSPIDER available online for converting population genetic data which enable handling microsatellite data.

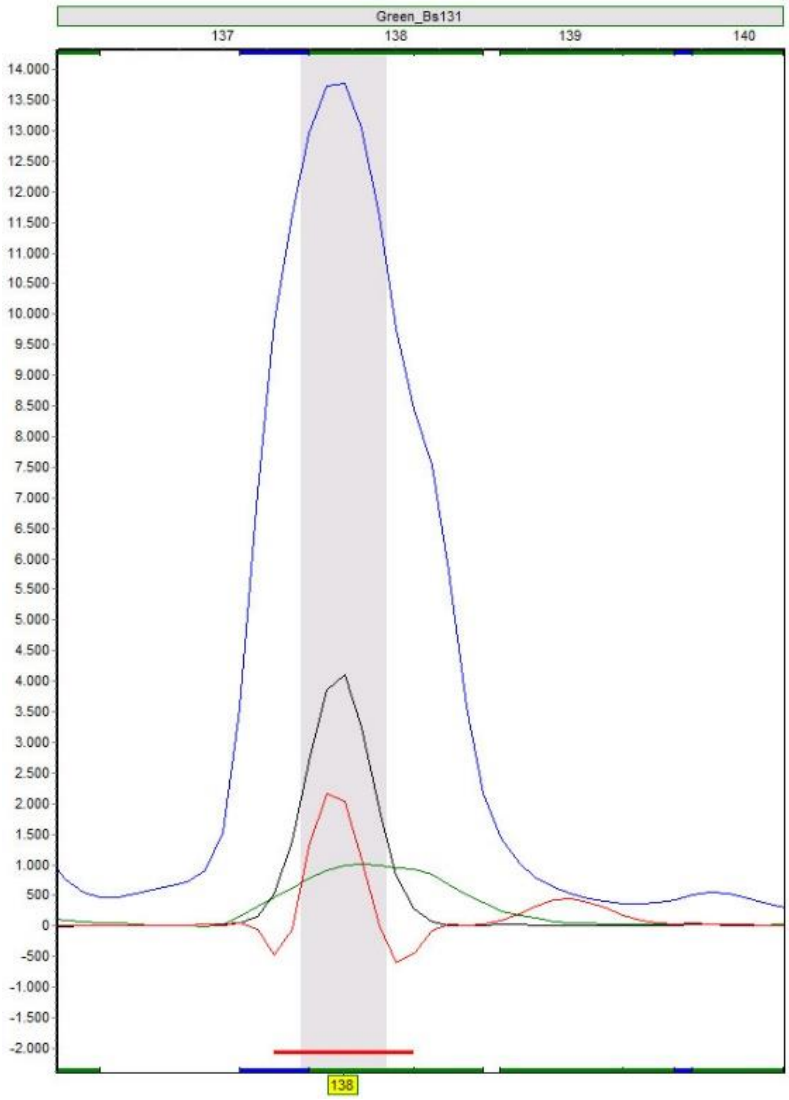


**Figure 1.** Pull up caused by overlap of fluorophores.

**Convert**

CONVERT is a simple program that converts files containing diploid allelic data. CONVERT reads GENEPOP input format and its own format (*.tab delimited text) which can be easily produced from excel file and converts input data files for GENEPOP, GDA (*.nex), ARLEQUIN (*.arp), POPGENE (*.dat), MICROSAT(*.mst), PHYLIP and STRUCTURE (*.str). Besides, CONVERT also produces allele frequency tables for each locus which allows a user to visually compare allele frequencies across populations (Glaubitz, 2004).

**Formatomatic**

FORMATOMATIC generates input files for ARLEQUIN, GENEPOP, MSVAR, BAPS, IMMANC/BAYESASS+, MIGRATE, NEWHYBRIDS, STRUCTURE, TM^/CONE, TMVP, MLNE, MRATIO, from *.pnr, genepop or csv write format (Excel) files. Likewise, it only processes diploid microsatellite data (Manoukis, 2007).

**Genotyping Error Checking**

Low quality of template DNA, mutations at primer binding sites, errors in PCR reaction may lead to incorrect assignment of microsatellite genotypes. Genotyping errors may occur due to the low DNA concentration which may result in failure of amplifying allele (allelic dropout) (Miller & Waits, 2003) and amplification of small alleles (large allele dropout) or in other words failure of large allele amplification (Wattier, Engel, Saumitou-Laprade, & Valero, 1998). Errors in PCR amplification can produce additional peaks that are very common in dinucleotide loci and makes it difficult to distinguish heterozygotes and homozygotes. False homozygotes may be generated when a mutation occurs at primer binding site (null alleles)(Shaw, Pierce, & Boyle,

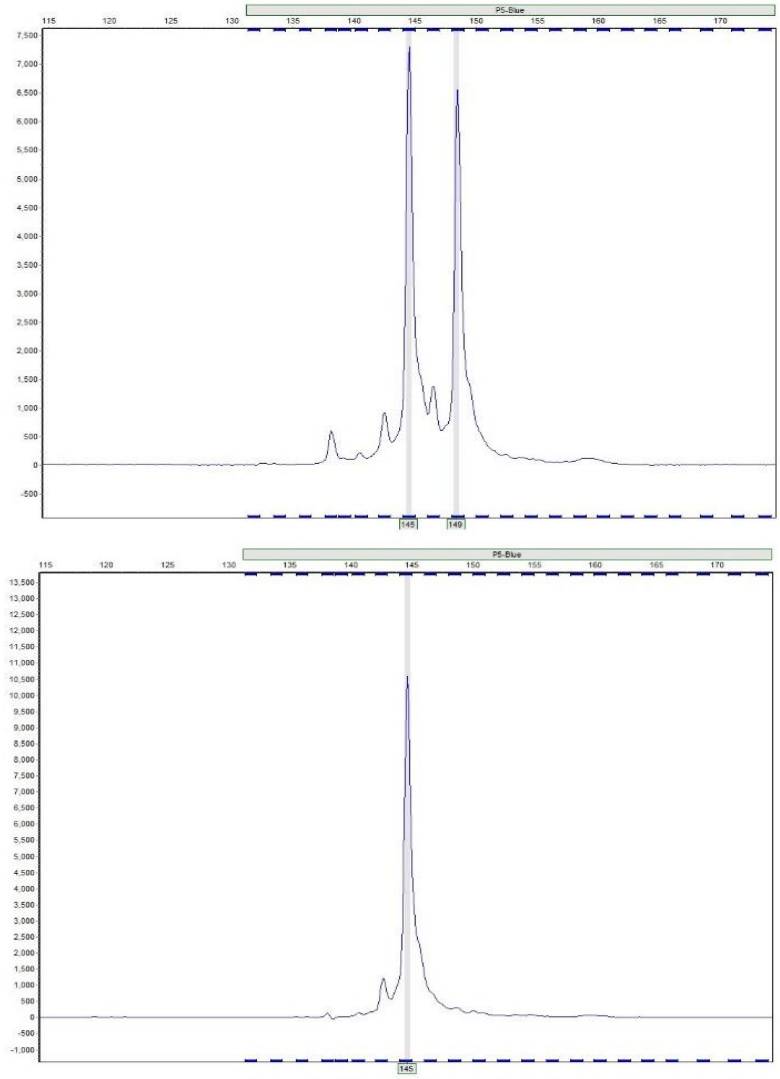

**Figure 2.** Heterozygote with closely sized alleles (upper), allele dropout; allele at a locus that consistently fails to amplify (lower).

1999). Such errors can cause deviations from Hardy-Weinberg equilibrium. MICROCHECKER (Van Oosterhout, Hutchinson, Wills, & Shipley, 2004), PEDMANAGER (Ewen *et al*., 2000), CERVUS (Marshall, Slate, Kruuk, & Pemberton, 1998) and GIMLET (Valiere, 2002) are among well-known programs that offer genotyping error detection.

**Microchecker**

MICROCHECKER, developed by Van Oosterhout *et al*. (2004), detects and identifies genotyping errors resulting from null alleles, allele dropout and stuttering. It can also distinguish inbreeding, Wahlund effects and deviation from Hardy-Weinberg equilibrium if multi-locus genotypes are available. Furthermore, null allele frequency can be estimated if null alleles are present. MICROCHECKER takes different data input formats including Microsoft Excel, text and GENEPOP.

**Calculation of Population Parameters**

After sorting, formatting and error-checking processes of raw genotype file data, one can proceed with calculation of basic population parameters. There are numerous software programs but GENEPOP, FSTAT, GenAIEx, Arlequin and Cervus are among the most popular programs that offer options for calculating genetic diversity, gene flow, F-statistics and so on.

**Genepop**

GENEPOP is a basic population genetic software developed by Raymond and Rousset (1995). The latest version of GENEPOP 4.6. (Rousset, 2008) run on R Package, Linux, Mac OCX and Microsoft platforms. GENEPOP input file can be converted from Excel file by using CONVERT. Each allele is coded by two numbers, population number or loci is

not limited. Moreover, missing data can be handled. GENEPOP displays a simple menu, which enables population data analyses such as; Hardy-Weinberg test, genotypic disequilibrium, an exact test for population differentiation, private allele method for NM estimate, allele frequencies, F-statistics and gene diversities.

**Fstat**

FSTAT, tests and estimates gene diversities and F-statistics. FSTAT harnesses randomization method to test the data. FSTAT converts GENEPOP format to FSTAT (*.dat) and vice-versa. FSTAT is a user-friendly software that has an easy access interface. FSTAT calculates allele frequency (per sample and overall), genotypic frequency (per sample and locus), gene diversity (per sample and locus), number of alleles (per sample, locus and overall), allelic richness (per locus, sample and overall) and Fis (per locus and sample). Furthermore, FSTAT tests Hardy-Weinberg equilibrium, genotypic equilibrium and also performs partial Mantel test and multiple regression. (Goudet, 1995)

**GenAIEx**

GenAIEx (Genetic Analysis in Excel) is an add-in for Microsoft office which provides additional features for population genetic analysis of codominant, haploid and binary genetic data including microsatellite. GenAIEx offers numerous graphical options for the output of genetic analysis, all of which can be modified by a user with Excel options. Beside basic population genetic statistics such as number of alleles, observed and expected heterozygosity and allele frequencies, GenAIEx offers a vast range of analyses including Nei's genetic distance measures, F-Statistics ($F_{ST}$, $F_{IS}$, $F_{IT}$), genotypic probability estimation, population assignment etc. Furthermore, GenAIEx performs a calculation of genetic and geographic distance matrix, which enables to perform Mantel test, spatial auto correlation and principal coordinate analysis (Peakall & Smouse, 2006, 2012).

**Arlequin**

Arlequin is a package program which performs inter and intra population analysis including molecular diversity, linkage disequilibrium, Fu's neutrality test, Mantel test, AMOVA, $F_{ST}$-pairwise genetic distances, pairwise molecular distances and many more. New version was recently released (Arlequin 3.5) which run under both Linux and Windows. While older version produces output files in HTML format, new version produces output files in XML format that can be parsed by R function to create graphical outputs. Arlequin is especially useful to examine genetic variation by hierarchical Analysis of Molecular Variance (AMOVA) by locality and sampling year within each locality (Excoffier & Lischer, 2010). The significance of molecular variance and indices is assessed with permutations of genotypes. Detailed information and implementation methods are provided on official web site.

**Cervus**

Cervus analyses genetic data generated from microsatellites and SNPs. Cervus operates on two assumptions; i) Markers are in linkage equilibrium, ii) markers are autosomal and species is diploid. While Cervus mainly used for parentage analysis it also offers many features such as file conversion (converts genotype file from or to Cervus format), genotyping error checking, allele frequency analysis and identity analysis (Kalinowski, Taper, & Marshall, 2007). Cervus is able to handle thousands of loci within a dataset and most of the bugs are fixed with the newest version (3.0.7).

**Advanced Analysis**

This section is for those who require additional population analyses to characterize subpopulations, to test population demography, isolation by distance, population size and so on. Programs such as STRUCTURE, Bottleneck, GenAlEx, LDNe, SAMOVA and Migrate can be used in advanced analysis.

**Structure**

STRUCTURE is a software package for multi-locus genotype data (Microsatellite, SNP, RFLPs and AFLPs) to investigate population structure (Pritchard, Stephens, & Donnelly, 2000). It can detect the presence of distinct populations, hybrid zones, migrants and admixed individuals. STRUCTURE operates by minimizing linkage disequilibrium, not by minimizing departure from linkage disequilibrium. Thus, you can either say "by minimizing linkage disequilibrium" or "by minimizing departure from linkage equilibrium". Identification of most likely K value is performed by Monte Carlo searches for the putative clusters (k, from 2 to …). Monte Carlo searches are generally run for $1x10^8$ steps with $10^7$ burning steps for all K groups with 3 independent runs for each K. STRUCTURE performs well while inferring population structure even if at low levels of genetic differentiation between subpopulations ($0.02<F_{ST}<0.1$) (Putman & Carbone, 2014).

The output of the structure is a basic plotting and reporting of results. Better graphical displays of structure can be produced by CLUMPP (Jakobsson & Rosenberg, 2007) and DISTRUCT (Rosenberg, 2004). Additionally, STRUCTURE HARVESTER, online program, offers additional visuals for STRUCTURE output and detects most likely K value. It takes zipped (using Winzip) result files of

STRUCTURE as an input file. It runs online and available at http://taylor0.biology.ucla.edu/structureHarvester/

### Tess

TESS v.2.3.1 analyses spatial population genetics with Bayesian clustering and Markov models (Chen, Durand, Forbes, & Francois, 2007). The program performs individual geographical assignment and uses same data format with Structure (*.str). TESS is especially useful for detecting genetic discontinuities and estimating spatial variation. TESS also offers a graphical output of geographical clusters and visualize admixture proportions. TESS runs on Windows, Mac and Linux.

### Samova

SAMOVA 2.0 is a program that assesses the genetic structure of population by defining groups of population, which are differentiated from each other without constraint for the geographic information. It needs two input files; i) geographic coordinates of the sampling localities (*.geo) and ii) Arlequin input file (*.arp). Voronoi polygons are generated from sampling locations. Assignment of populations into K groups is random. Genetic barriers between groups are clarified as edges of Voronoi polygons begin to separate populations. Identification of genetic barriers is a by-product. SAMOVA 2.0 only runs on Windows (Dupanloup, Schneider, & Excoffier, 2002).

### Bottleneck

BOTTLENECK 1.2 is a software package that identifies populations, which have experienced a severe reduction in population size. The program runs on Windows platform and requires allele frequency data from at least 20 individuals and at least four polymorphic loci. BOTTLENECK uses GENEPOP format, conducts three tests; i) sign test, ii) standardized differences test, iii) Wilcoxon sign rank test with three Mutation models which are infinite allele model (IAM), stepwise mutation model (SMM) and two-phase model (TPM). Generally, TPM model is selected in microsatellite analysis. In order to get higher statistical power, at least 10 polymorphic microsatellite loci and sampling at least 30 individuals per sampling area is suggested (Piry, Luikart, & Cornuet, 1999).

### LDNe

LDNe is a program for estimation of effective population size based on linkage disequilibrium.

The program uses Genepop or FSTAT input data. There is no limitation in terms of a number of individual, loci and alleles (Waples & Do, 2008). LDne uses the Burrow method that does not depend on random mating assumption and does not require any haplotypic frequency information.

## Conclusion

The information about different software programs for population genetic data analysis, suitable for most of the eukaryotes including aquatic organisms, was summarized. Steps of microsatellite data analysis are linked with each other. Accuracy at initial steps; genotyping, scoring and error checking is the key to get reliable results. Any error at genotyping might change the outcome dramatically. Software programs, which are well accepted by the majority of users, were mentioned in this review. In general, it is advisable to use more than one software programs to compare the outputs for any parameter if any alternative is available. Software programs for calculation of population parameters and advanced analysis are generally easy to handle. Most of them are supported with online manuals which are open access and available to be downloaded at websites.

## References


Antoni, L., Luque, P. L., Naghshpour, K., & Saillant, E. A. (2014). Polymorphic microsatellite markers for the Yellowfin tuna (*Thunnus albacares*). *Conservation Genetics Resources, 6*(3), 609-611. doi:10.1007/s12686-014-0152-5

Baillie, S. M., Muir, A. M., Scribner, K., Bentzen, P., & Krueger, C. C. (2016). Loss of genetic diversity and reduction of genetic distance among lake trout *Salvelinus namaycush* ecomorphs, Lake Superior 1959 to 2013. *Journal of Great Lakes Research, 42*(2), 204-216. doi:10.1016/j.jglr.2016.02.001

Castoe, T. A., Poole, A. W., de Koning, A. P. J., Jones, K. L., Tomback, D. F., Oyler-McCance, S. J., . . . Pollock, D. D. (2012). Rapid Microsatellite Identification from Illumina Paired-End Genomic Sequencing in Two Birds and a Snake. *Plos One, 7*(2), e30953. doi:10.1371/journal.pone.0030953

Chen, C., Durand, E., Forbes, F., & Francois, O. (2007). Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes, 7*(5), 747-756. doi:10.1111/j.1471-8286.2007.01769.x

De Vasconcellos, A. V., Lima, D., Bonhomme, F., Vianna, M., & Sole-Cava, A. M. (2015). Genetic population structure of the commercially most important demersal fish in the Southwest Atlantic: The whitemouth croaker (*Micropogonias furnieri*). *Fisheries Research, 167*, 333-337. doi:10.1016/j.fishres.2015.03.008

Dupanloup, I., Schneider, S., & Excoffier, L. (2002). A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology, 11*(12), 2571-2581. doi:10.1046/j.1365-294X.2002.01650.x

Ellis, J. S., Gilbey, J., Armstrong, A., Balstad, T., Cauwelier, E., Cherbonnel, C., . . . Stevens, J. R. (2011). Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (*Salmo salar* L.). *Genetica, 139*(3), 353-367.

doi:10.1007/s10709-011-9554-4

Ewen, K. R., Bahlo, M., Treloar, S. A., Levinson, D. F., Mowry, B., Barlow, J. W., & Foote, S. J. (2000). Identification and analysis of error types in high-throughput genotyping. *American Journal of Human Genetics, 67*(3), 727-736. doi:10.1086/303048

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources, 10*(3), 564-567. doi:10.1111/j.1755-0998.2010.02847.x

Gemayel, R., Cho, J., Boeynaems, S., & Verstrepen, K. J. (2012). Beyond Junk-Variable Tandem Repeats as Facilitators of Rapid Evolution of Regulatory and Coding Sequences. *Genes, 3*(3), 461-480. doi:10.3390/genes3030461

Glaubitz, J. C. (2004). CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes, 4*(2), 309-310. doi:10.1111/j.1471-8286.2004.00597.x

Goudet, J. (1995). FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity, 86*(6), 485-486.

Harrison, H. B., Feldheim, K. A., Jones, G. P., Mansour, H., Perumal, S., Williamson, D. H., & Berumen, M. L. (2014). Validation of microsatellite multiplexes for parentage analysis in a coral reef fish (*Lutjanus carponotatus*, *Lutjanidae*). *Conservation Genetics Resources, 6*(4), 803-806. doi:10.1007/s12686-014-0226-4

Hauser, L., Adcock, G. J., Smith, P. J., Ramirez, J. H. B., & Carvalho, G. R. (2002). Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences of the United States of America, 99*(18), 11742-11747. doi:10.1073/pnas.172242899

Hillen, J. E. J., Coscia, I., Vandeputte, M., Herten, K., Hellemans, B., Maroso, F., . . . Volckaert, F. A. M. (2017). Estimates of genetic variability and inbreeding in experimentally selected populations of European sea bass. *Aquaculture, 479*, 742-749. doi:10.1016/j.aquaculture.2017.07.012

Hoffman, J. I., & Amos, W. (2005). Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology, 14*(2), 599-612. doi:10.1111/j.1365-294X.2004.02419.x

Holleley, C.E., & Geerts, P.G. (2009). Multiplex Manager 1,0: a cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques*, 46(7), 511-517

Hollenbeck, C. M., Portnoy, D. S., & Gold, J. R. (2015). A genetic linkage map of red drum (Sciaenops ocellatus) and comparison of chromosomal syntenies with four other fish species. *Aquaculture, 435*, 265-274. doi:10.1016/j.aquaculture.2014.08.045

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics, 23*(14), 1801-1806. doi:10.1093/bioinformatics/btm233

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology, 16*(5), 1099-1106. doi:10.1111/j.1365-294X.2007.03089.x

Li, Y., Zhao, H., Peng, Z., & Zhang, Y. (2016). Isolation and characterization of 21 polymorphic microsatellite markers for a new labeonine fish (*Paraqianlabeo lineatus* Zhao *et al*. 2014) using illumina paired-end sequencing. *Journal of Applied Ichthyology, 32*(1), 126-128. doi:10.1111/jai.12938

Lv, W. H., Zheng, X. H., Kuang, Y. Y., Cao, D. C., Yan, Y. Q., & Sun, X. W. (2016). QTL variations for growth-related traits in eight distinct families of common carp (*Cyprinus carpio*). *BMC Genetics, 17*, 65. doi:10.1186/s12863-016-0370-9

Manoukis, N. C. (2007). FORMATOMATIC: a program for converting diploid allelic data between common formats for population genetic analysis. *Molecular Ecology Notes, 7*(4), 592-593. doi:10.1111/j.1471-8286.2007.01784.x

Marshall, T. C., Slate, J., Kruuk, L. E. B., & Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology, 7*(5), 639-655. doi:10.1046/j.1365-294x.1998.00374.x

McKeown, N. J., Hauser, L., & Shaw, P. W. (2017). Microsatellite genotyping of brown crab *Cancer pagurus* reveals fine scale selection and 'non-chaotic' genetic patchiness within a high gene flow system. *Marine Ecology Progress Series, 566*, 91-103. doi:10.3354/meps12044

Miller, C. R., & Waits, L. P. (2003). The history of effective population size and genetic diversity in the Yellowstone grizzly (Ursus arctos): Implications for conservation. *Proceedings of the National Academy of Sciences of the United States of America, 100*(7), 4334-4339. doi:10.1073/pnas.0735531100

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes, 6*(1), 288-295. doi:10.1111/j.1471-8286.2005.01155.x

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics, 28*(19), 2537-2539. doi:10.1093/bioinformatics/bts460

Phumichai, C., Phumichai, T., & Wongkaew, A. (2015). Novel Chloroplast Microsatellite (cpSSR) Markers for Genetic Diversity Assessment of Cultivated and Wild Hevea Rubber. *Plant Molecular Biology Reporter, 33*(5), 1486-1498. doi:10.1007/s11105-014-0850-x

Piry, S., Luikart, G., & Cornuet, J. M. (1999). BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity, 90*(4), 502-503. doi:10.1093/jhered/90.4.502

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution, 4*(22), 4399-4428. doi:10.1002/ece3.1305

Raymond, M., & Rousset, F. (1995). Genepop (Version-1.2) - Population-Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity, 86*(3), 248-249.

Renshaw, M. A., Saillant, E., Bradfield, S. C., & Gold, J. R. (2006). Microsatellite multiplex panels for genetic studies of three species of marine fishes: red drum (*Sciaenops ocellatus*), red snapper (*Lutjanus campechanus*), and cobia (*Rachycentron canadum*).

*Aquaculture*, *253*(1-4), 731-735. doi:10.1016/j.aquaculture.2005.09.012

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, *4*(1), 137-138. doi:10.1046/j.1471-8286.2003.00566.x

Rousset, F. (2008). GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103-106. doi:10.1111/j.1471-8286.2007.01931.x

Shaw, P. W., Pierce, G. J., & Boyle, P. R. (1999). Subtle population structuring within a highly vagile marine invertebrate, the veined squid Loligo forbesi, demonstrated with microsatellite DNA markers. *Molecular Ecology*, *8*(3), 407-417. doi:10.1046/j.1365-294X.1999.00588.x

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology & Evolution*, *15*(5), 199-203. doi:10.1016/S0169-5347(00)01825-5

Valiere, N. (2002). GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes*, *2*(3), 377-379. doi:10.1046/j.1471-8286.2002.00228.x

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), 535-538. doi:10.1111/j.1471-8286.2004.00684.x

Waples, R. S., & Do, C. (2008). LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, *8*(4), 753-756. doi:10.1111/j.1755-0998.2007.02061.x

Wattier, R., Engel, C. R., Saumitou-Laprade, P., & Valero, M. (1998). Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology*, *7*(11), 1569-1573. doi:10.1046/j.1365-294x.1998.00477.x