Journal of Heredity, 2025, **116**, 34–42
https://doi.org/10.1093/jhered/esae060
Advance access publication 22 October 2024
Original Article

Original Article

ParthenoGenius: A user-friendly heuristic for inferring presence and mechanism of facultative parthenogenesis from genetic and genomic datasets

Brenna A. Levine[1,2,*], **Warren Booth**[2,3]

[1]Department of Biology, Kean University, Union, NJ, United States,
[2]Chiricahua Desert Museum, Rodeo, NM, United States,
[3]Department of Entomology, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

[*]Corresponding author: Department of Biology, Kean University, 1000 Morris Ave, Union, NJ 07083, United States. Email: brlevine@kean.edu

Corresponding Editor: Oliver Ryder

Abstract

Facultative parthenogenesis (FP), or asexual reproduction by sexually reproducing female animals, has been reported across several clades of vertebrates and is increasingly being recognized as a reproductive mechanism with significant implications for the genetic variation of captive and wild populations. The definitive identification of parthenogens requires molecular confirmation, with large genomic datasets necessary to accurately parse the parthenogenetic mechanism (i.e. endoduplication, gametic duplication, terminal fusion automixis, or central fusion automixis). Current methods for inferring FP from large genomic datasets are statistically intensive, require competency in R scripting for their execution, and are not designed for detection of facultative parthenogenesis or screening of large numbers of mother/offspring pairs, whereas small datasets (i.e. microsatellites) that can be evaluated visually lack the power to discriminate among FP mechanisms. Here, we present the user-friendly software program, *ParthenoGenius*, that uses intuitive logic to infer the presence and mechanism of FP from even large genomic datasets comprising many mothers and offspring. *ParthenoGenius* runs relatively quickly and does not require the researcher to have knowledge of R scripting or statistics. *ParthenoGenius* was tested on eight empirical datasets and in each case identified parthenogens (and parthenogenic mechanism when present) consistent with results of previous studies or corroborating evidence. *ParthenoGenius* will facilitate the rapid screening of large genomic datasets comprising many mothers and offspring for the presence and mechanism of parthenogenesis, improving our understanding of the frequency and phylogenetic distribution of FP across the animal kingdom.

Key words: automixis, central fusion, endoduplication, gametic duplication, sperm storage, terminal fusion

Introduction

Facultative parthenogenesis (FP) (i.e. sporadic asexual reproduction by sexually reproducing species) has been reported across four major taxa in the animal kingdom: namely, crocodiles0020(Booth et al. 2023), reptiles (lizards and snakes) (Schuett et al. 1997; Watts et al. 2006; Germano and Smith 2010; Booth and Schuett 2011, 2016; Booth et al. 2011a, 2011b, 2012, 2014; Reynolds et al. 2012; Kinney et al. 2012; Shibata et al. 2017; Allen et al. 2018; Miller et al. 2019; Kratochvíl et al. 2020; Seixas et al. 2020; Card et al. 2021), birds (Olson 1960; Schut et al. 2008; Parker and McDaniel 2009; Ramachandran and McDaniel 2018; Ryder et al. 2021), and elasmobranch fishes (Chapman et al. 2007, 2008; Feldheim et al. 2010; Robinson et al. 2011; Portnoy et al. 2014; Fields et al. 2015; Harmon et al. 2016; Straube et al. 2016). Historically, vertebrate FP was considered to be a reproductive syndrome of captivity, with cases restricted to zoological collections and private breeding facilities. However, with its detection in natural populations (e.g. copperhead and cottonmouth snakes [Booth et al. 2012]; smalltooth sawfish [Fields et al. 2015]), the ecological and evolutionary significance of this reproductive trait is now being questioned, notably for the potentially significant implications it may have for population genetic variation.

In vertebrates, two cytological mechanisms of FP are known—Endoduplication and automixis. The effects of these range from the complete retention of maternal heterozygosity to genome-wide homozygosity in the offspring (Lampert 2008). Endoduplication involves the premeiotic doubling of chromosomes and the production of diploid eggs. Recombination takes place but is masked as this occurs between homologous chromosomes; thus, the offspring will be genetically identical to the mother, retaining 100% of her heterozygosity. In contrast, under automixis, meiosis is maintained, and diploidy is recovered by one of three mechanisms: 1) gametic duplication, which will result in genome-wide homozygosity, 2) terminal fusion, resulting from the fusion of the nucleus with the second polar body and hence the potential to retain some maternal heterozygosity (0% to 67%) largely localized at the chromosome

Received February 16, 2024; Accepted October 18, 2024

tips, and 3) central fusion automixis, in which the nucleus fuses with the first polar body, leading to the retention of significant maternal heterozygosity (67% to 100%) (Pearcy et al. 2006, 2011; Svendsen et al. 2015). Of these four potential mechanisms, terminal fusion appears to be the most commonly attributed mechanism in vertebrates (Booth and Schuett 2016).

From a practical standpoint, the recognition that FP can occur in wild populations poses a challenge for those conducting genetic parentage analysis and pedigree reconstruction. First, most genetic parentage analysis programs assume exclusive sexual reproduction or selfing (e.g. *Colony* [Jones and Wang 2010]; *Sequoia* [Huisman 2017]) with no consideration of inheritance patterns associated with FP, such that in populations that have FP, it is often assumed that an unsampled male sired progeny for which paternity could not be assigned. Furthermore, many female vertebrates are capable of long-term sperm storage (LTSS) (Holt and Loyd 2010; Booth and Schuett 2011; Levine et al. 2021, 2024), a phenomenon that may produce offspring long after a mating event, with those offspring mistakenly attributed to FP. To ensure accurate genetic parentage assignments and downstream analyses, the identification of parthenogenetic offspring should be prioritized within parentage analysis workflows for taxa that lack genomic imprinting (i.e. the mechanism that prevents FP in mammals [Kono 2006]) and thus for which FP may occur (Solter 1988).

Definitively parsing FP from LTSS and undetected mating events requires analysis of genetic or genomic data. Most often, microsatellite genotypes have been evaluated to infer FP. While microsatellite data are easily evaluated visually, the discriminatory power of these to parse FP mechanisms is low, potentially leading to erroneous inferences (e.g. Shibata et al. 2017). Large genomic datasets, on the other hand, capture many sites in the genome and are therefore more likely to sequence sites near the tips of chromosomes, thus allowing for the parsing of FP mechanism. Furthermore, in the era of next-generation sequencing, the generation of genomic parentage datasets is now commonplace (e.g. Kess et al. 2016; Andrews et al. 2018; Thrasher et al. 2018; Levine et al. 2019, 2022). Yet, to our knowledge, no tool currently exists to facilitate rapid screening of genomic parentage datasets for the presence of parthenogens or to discern the mechanism underlying apparent FP. Recent studies of FP that analyzed genomic datasets (i.e. Allen et al. 2018; Card et al. 2021) relied on statistically intensive methods designed for the study of inbreeding (e.g. *Rhh* [Alho et al. 2010]) or relatedness (e.g. *Demerelate* [Kraemer and Gerlach 2017]) rather than detection of FP. Further, these programs require the user to have competency in R scripting and are not optimized for screening large numbers of mothers and offspring for the presence of FP. This is unfortunate in that the ability to rapidly screen genomic parentage datasets comprising many mothers and offspring of species for which FP is possible has the potential to dramatically improve our understanding of the frequency and phylogenetic distribution of FP and LTSS in nature.

Here, we present a user-friendly heuristic, *ParthenoGenius*, that uses intuitive logic to quickly scan genetic and genomic data comprising even large numbers of mothers and offspring for the presence of FP. When FP is detected, this program also infers the most likely FP mechanism. *ParthenoGenius* will facilitate rapid screening of genetic parentage datasets for parthenogens, allowing for unprecedented characterization of rates of FP and LTSS in captive and wild populations and improving our understanding of the phylogenetic distribution of FP among vertebrates.

Methods

ParthenoGenius is an open-source software program written in Python that requires the *pandas* (https://pandas.pydata.org) module. The *ParthenoGenius* script (ParthenoGenius.py), a shell script for iteratively running *ParthenoGenius* on multiple mother–offspring pairs (PG_wrapper.sh), and a file preparation shell script (PG_file_prep.sh) that generates mother–offspring input files from genetic parentage datasets comprising many individuals can be found on GitHub (https://github.com/brenna-levine/LEVINE_ParthenoGenius) along with instructions for installation and usage.

Overview

ParthenoGenius employs heuristics for inferring the presence and mechanism of FP for diploid organisms that are based on expectations of the relationship between maternal and offspring genotypes when offspring are produced via FP. In a parthenogen, offspring will be homozygous for the maternal allele at all loci for which the mother is homozygous, as no paternal alleles are present. It follows that heterozygous or opposite homozygous offspring genotypes observed at maternal homozygous loci in a parthenogen are artifacts of sequencing and/or genotyping error (i.e. Mendelian errors [Fountain et al. 2016]), and to a nominal extent, germline mutation. While mutations may cause the offspring genotypes at maternal homozygous loci to differ from the mother, per-generation mutation rates are exceedingly low in vertebrates (e.g. 1.17×10^{-8} for reptiles and 1.01×10^{-8} for birds [Bergeron et al. 2023]) compared to sequencing and genotyping error rates estimated from Mendelian errors (e.g. ≥0.03 for ≥5x coverage reference-aligned ddRADseq [Fountain et al. 2016]) and are therefore not considered by *ParthenoGenius*. Using this logic, Part 1 of the *ParthenoGenius* program scans a genotype file containing maternal and offspring alleles, identifies all loci at which the mother is homozygous, and tallies the number of these loci at which the offspring does not display homozygosity for the maternal allele. If the number of maternal homozygous loci at which the offspring displays differing genotypes exceeds that expected from sequencing and genotyping error alone, *ParthenoGenius* calls the offspring as a nonparthenogen as it is assumed that the presence of paternal alleles is accounting for the observed differences (Fig. 1). In this case, *ParthenoGenius* generates two output files and shuts down: 1) A text file that lists all diploid co-dominant loci at which the mother is homozygous and the offspring alleles that differ from the maternal genotype at these loci, and 2) a summary file that reports metrics related to maternal and offspring homozygosity and a statement that the offspring is unlikely to be a parthenogen. Differently, if the number of maternal homozygous loci at which the offspring displays differing genotypes is less than that expected from sequencing and genotyping error alone, *ParthenoGenius* generates the above homozygosity output files, states that the offspring is likely a parthenogen in the summary file, and proceeds to Part 2 of the program.

Part 2 of *ParthenoGenius* similarly uses a heuristic to infer the reproductive mechanism responsible for FP based on expectations of offspring heterozygosity given different FP mechanisms. Here, *ParthenoGenius* identifies all loci at which

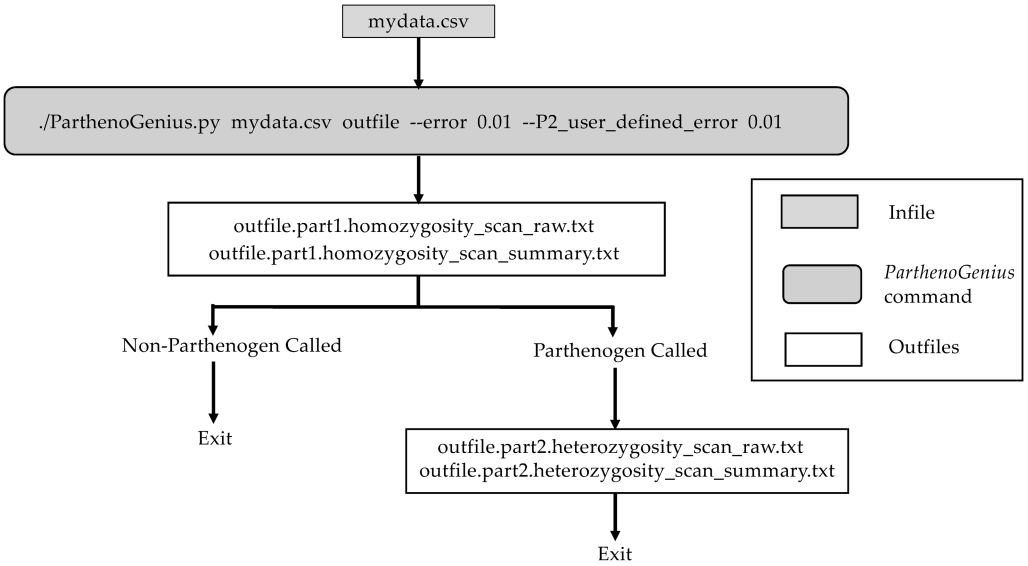

Fig. 1. Overview of *ParthenoGenius*. *ParthenoGenius* takes as an input file a.csv file containing locus names and genotypes for a mother and offspring (e.g. mydata.csv). The user must declare a prefix for the generated output files (e.g. "outfile") and can optionally declare an estimated error rate as an argument for Part 1 (--error; default = 0.001) and Part 2 (--P2_user_defined_error; default is calculated by Part 1). If the offspring is called as a non-parthenogen, two output files regarding offspring homozygosity at maternal homozygous loci are generated, and the program shuts down. If the offspring is called a parthenogen, two additional output files regarding offspring homozygosity at maternal heterozygous loci are generated, and the program calls the most likely mechanism of parthenogenesis. Detailed instructions for installation and usage of *ParthenoGenius* are found on GitHub (https://github.com/brenna-levine/LEVINE_ParthenoGenius).

the mother is heterozygous and tallies the number of these loci at which the offspring displays heterozygosity for the maternal alleles. If the number of maternal heterozygous loci at which the offspring is heterozygous is less than would be expected due to the error rate alone assuming gametic duplication (i.e. genome-wide offspring homozygosity), the mechanism of parthenogenesis is called as gametic duplication. If the number is greater than would be expected for gametic duplication but less than would be expected due to central fusion (i.e. 67% or greater; Svendsen et al. 2015), the mechanism of parthenogenesis is called as terminal fusion. If the number exceeds the minimum expected due to central fusion automixis but is less than the number expected due to the error rate alone assuming endoduplication (i.e. <100%), the mechanism of parthenogenesis is called as central fusion automixis. If the offspring is heterozygous at more loci than would be expected due to the error rate alone assuming endoduplication, the mechanism is called as endoduplication. Note that central fusion automixis can also theoretically yield 100% retained heterozygosity, but this is extremely unlikely as it would only happen in the absence of recombination. Likewise, while 0% retained maternal heterozygosity is theoretically possible for terminal fusion automixis, this result is extremely unlikely as it would also require the absence of recombination. Part 2 of *ParthenoGenius* generates two output files: 1) A file that lists all loci at which the mother is heterozygous and corresponding heterozygous offspring alleles and 2) a summary file that reports heterozygosity values and a statement of the most likely mechanism of FP.

Error rates

A critical parameter for *ParthenoGenius* is the estimated error rate for the genetic dataset. The value of this parameter is used in Part 1 of *ParthenoGenius* to generate expectations for the maximum number of maternal homozygous loci at which the offspring should exhibit heterozygosity or opposite homozygosity, given the offspring is a parthenogen. The default error rate for Part 1 of *ParthenoGenius* is 0.001 (0.1%), a value consistent with the error rate expected from a high-quality SNP array (Saunders et al. 2007). However, the error rate can be provided as an argument by the user to reflect the DNA sequencing technology, coverage, DNA quality, base call quality, and locus assembly protocol (e.g. reference-aligned vs. de novo) for other sequencing methods such as reduced representation library sequencing. For example, short read Illumina error rates have been estimated at 0.1% to 1.0% (Peterson et al. 2012), and variable coverage (e.g. ≥5x vs. ≥30x) has resulted in an error rate ranging from ≥0.03 to ≤0.01, respectively, in reference-aligned ddRADseq datasets (Fountain et al. 2016). The user should carefully consider their choice of error rate for part one of *ParthenoGenius* and test the robustness of their analyses to various values (see *Error rate tests* section).

In Part 2 of *ParthenoGenius*, the error rate is updated by default to reflect the error rate apparent among the homozygous loci, assuming the offspring is a parthenogen. The updated error rate is equal to the number of maternal homozygous loci at which the offspring displays heterozygosity or opposite homozygosity divided by the total number of maternal homozygous loci. This updated error rate assumes that the error rate is equivalent among homozygous and heterozygous loci, which may or may not be a reasonable assumption for the user's dataset (e.g. depending on coverage). Therefore, Part 2 of *ParthenoGenius* can also be executed using a user-defined error rate that is not informed by the error rate apparent among the homozygous loci to accommodate potential differences in error rate among heterozygous and homozygous loci.

Analyses of simulated datasets

We simulated five test datasets by hand reflecting nonparthenogenesis and FP due to gametic duplication, terminal fusion automixis, central fusion automixis, and endoduplication. Each simulated dataset contained 100 SNP loci and genotypes for a simulated mother and offspring. One hundred SNPs were chosen to represent a compromise between large SNP datasets and small microsatellite datasets. In all datasets, the mother was homozygous at 36 loci, but only 33 loci were genotyped for both mother and offspring. Missing data was included in the simulated datasets to reflect the realities of empirical data. A summary of each simulated dataset is available in Supplementary Table 1. We ran *ParthenoGenius* on each simulated dataset using an error rate equal to 0.01 (1%), reflecting a potential error rate for high-coverage (≥30x), reference-aligned ddRADseq loci (Fountain et al. 2016).

Analyses of empirical datasets

We analyzed six published genetic and genomic empirical datasets to determine whether *ParthenoGenius* generated conclusions consistent with those previously reported. We also used *ParthenoGenius* to analyze two novel empirical data (i.e. Booth et al. 2023; Levine et al. 2024). We set the per-base error rate equal to 0.01 (1%) for microsatellite data (i.e. a value that has been previously used for microsatellite parentage analysis simulations [Kalinowski et al. 2007]) and ddRADseq data and 0.001 (0.1%) for a high-coverage whole genome sequencing (WGS) dataset.

Published dataset 1: Green anaconda FP. Shibata et al. (2017) used 16 microsatellite loci to analyze the DNA of two putatively parthenogenetic green anacondas (*Eunectes murinus*) as well as two nonparthenogenetic individuals. The study concluded that the two candidate parthenogens were true parthenogens produced via gametic duplication due to homozygosity for maternal alleles at all loci.

Published dataset 2: California condor FP. Ryder et al. (2021) used 22 microsatellite loci to analyze the DNA of two putatively parthenogenetic California condors (*Gymnogyps californianus*) and concluded that both offspring were parthenogens due to homozygosity for maternal alleles at all loci. This study refrained from definitively calling the mechanism of FP due to the limited power of the microsatellite dataset.

Published dataset 3: Zebra shark FP. Dudgeon et al. (2017) used 14 microsatellite loci to analyze the DNA of 12 zebra shark (*Stegostoma fasciatum*) offspring and their mothers. The study identified nine parthenogenetic offspring, with the limited power of the microsatellite dataset implicating gametic duplication due to homozygosity for maternal alleles observed at all loci. However, the authors concluded that terminal fusion automixis was a more likely mechanism of FP for this species.

Published dataset 4: Western diamond-backed rattlesnake LTSS. Levine et al. (2021) used ddRADseq to genotype a western diamond-backed rattlesnake mother (*Crotalus atrox*) and four candidate parthenogens at 4,228 SNPs. The mother was homozygous at 1,197 loci, and all offspring exhibited putative paternal alleles at 5.1% of these loci. The authors rejected the hypothesis of FP in favor of LTSS due to the identification of putative paternal alleles and a parentage analysis using *COLONY* (Jones and Wang 2010).

Published dataset 5: Guatemalan beaded lizard multiple paternity. Levine et al. (2022) used ddRADseq to genotype a captive zoo population of Guatemalan beaded lizard (*Heloderma charlesbogerti*) mothers, candidate fathers, and 30 offspring at 3,094 SNPs. The authors conducted pedigree reconstruction using *Sequoia* (Huisman 2017) with 210 high minor allele frequency SNPs to identify incidents of multiple paternity in this captive population. All offspring were assigned paternity to a known male in the population.

Published dataset 6: King cobra FP. Card et al. (2021) used ddRADseq to genotype a king cobra mother (*Ophiophagus hannah*) and two candidate parthenogens at 20,562 SNPs. The mother was heterozygous at 278 loci. The study used the R programs *Rhh* (Alho et al. 2010) and *Demerelate* (Kraemer and Gerlach 2017) and a custom R script to perform bootstrapping to quantify measures of genome-wide heterozygosity and relatedness and concluded that the two offspring exhibited retained heterozygosity at 23% and 27% of the maternal heterozygous loci, respectively. Combined with observations of clutch characteristics, the study concluded that the two offspring were parthenogens produced via terminal fusion automixis.

Novel dataset 1: American crocodile FP. We used *ParthenoGenius* to analyze 191,255 SNPs genotyped by WGS for an American crocodile (*Crocodylus acutus*) mother and candidate parthenogen (Booth et al. 2023). FP via terminal fusion automixis was inferred from characteristics of the clutch, >99.9% homozygosity at 186,240 maternal homozygous loci, and 3% retained maternal heterozygosity in the offspring at 5,015 maternal heterozygous loci. Two differently thinned SNP datasets were also analyzed and generated the same conclusions.

Novel dataset 2: Himalayan mountain pitviper LTSS. We used *ParthenoGenius* to analyze 9,697 SNPs genotyped by ddRADseq for a Himalayan mountain pitviper (*Ovophis monticola*) mother and three candidate parthenogens (Levine et al. 2024). The offspring displayed putative paternal alleles at 31% to 39% of 4,087 maternal homozygous loci. No evidence of FP was found, instead implicating LTSS as the reproductive mechanism responsible for the offspring.

Error rate tests

We also ran *ParthenoGenius* on all of the described empirical datasets using seven user-defined error rates (0.001, 0.005, 0.010, 0.015, 0.020, 0.025, 0.030) for both Parts 1 and 2 of the program to test the sensitivity of the results to the choice of error rate. For each test, we recorded whether the offspring were called as a parthenogen and if so, what mechanism was ascribed. These results were then compared to those previously reported in the literature.

Results

Simulated and published datasets

ParthenoGenius called the presence and mechanism of FP as expected from the five simulated test datasets. *ParthenoGenius*, also called offspring as parthenogens, following analyses of

the three published microsatellite datasets, consistent with the results of these studies (Table 1). *ParthenoGenius* called the mechanism of parthenogenesis as gametic duplication for all microsatellite datasets, also consistent with statements or suggestions in these studies. However, we caution that calls for gametic duplication may be due to a lack of power in the microsatellite datasets to detect alternative mechanisms of FP (Allen et al. 2018; Card et al. 2021; Booth et al. 2023).

ParthenoGenius, also called parthenogenesis or nonparthenogenesis, is consistent with results previously reported in all analyses of published ddRADseq SNP datasets (Table 1). *ParthenoGenius* rejected the parthenogenetic origin of *C. atrox* (Levine et al. 2021) and *H. charlesbogerti* (Levine et al. 2022) offspring and called two *O. hannah* offspring as parthenogens (Card et al. 2021) (Table 1). Moreover, *ParthenoGenius* called the mechanism of FP for the *O. hannah* offspring as terminal fusion automixis due to the retention of 23% to 27% of maternal heterozygosity (Table 1), consistent with the findings of the original study. *ParthenoGenius* analyzed these large SNP datasets quickly on a personal laptop (2020 MacBook Pro with 2 GHz Quad-Core Intel Core-5 Processor with a solid-state drive). For example, 2 *O. hannah* offspring were analyzed for the presence and mechanism of parthenogenesis with the *ParthenoGenius* wrapper script in under 9 s despite each dataset consisting of 20,562 loci (Table 2). Similarly, *ParthenoGenius* screened 30 *H. charlesbogerti* offspring, each genotyped at more than 3,000 loci, for parthenogenesis in 30.28 s.

Of note, *ParthenoGenius* called *H. charlesbogerti* offspring as nonparthenogens despite some offspring exhibiting high homozygosity (i.e. 95.2%) at maternal homozygous loci (Table 1). This species is critically endangered, and samples originated from a small long-term breeding colony; therefore, high genome-wide homozygosity is not unexpected. We emphasize that *ParthenoGenius* assigned even those offspring with high homozygosity as nonparthenogens, consistent with the parentage analysis results of the previous study, indicating that assessment of FP with *ParthenoGenius* may be possible even for samples that are genetically depauperate.

Novel datasets

ParthenoGenius identified the first case of FP in a crocodilian, a finding which we published in Booth et al. (2023). The mechanism of FP was called as terminal fusion automixis, consistent with that of *O. hannah* (Card et al. 2021), due to 99.9% offspring homozygosity at maternal homozygous loci and retention of 3.0% of maternal heterozygosity (Table 1). These results were supported by observations of the clutch and knowledge of the mother's isolation from potential mates. *ParthenoGenius* also identified the first case of LTSS in a Himalayan mountain pitviper (*O. monticola*) by rejecting the hypothesis of FP for three male offspring due to the detection of contrasting offspring genotypes at 31% to 39% of maternal homozygous loci (Levine et al. 2024) (Table 1).

Error rate tests

Overall, *ParthenoGenius* was robust to the user-defined sequencing error rate for both Parts 1 and 2 of the program. In Part 1, nonparthenogens were classified as such regardless of the error rate used (Supplementary Table 2). Similarly, parthenogens were classified, and their mechanism ascribed to what was previously reported in the literature at nearly all error rates for both Parts 1 and 2. The only exception is with regard to the king cobra dataset. At an error rate of 0.001, both parthenogenetic offspring were called as nonparthenogens, inconsistent with previously published results. Similarly, at an error rate of 0.005, only one parthenogen was inferred

Table 1. We analyzed six published and two novel (Booth et al. 2023; Levine et al. 2024) datasets with *ParthenoGenius* to test for the presence and mechanism of FP.

Study	# Offspring	# Maternal homozygous loci	% Homozygosity at maternal homozygous loci in nonparthenogens	% Homozygosity at maternal homozygous loci in parthenogens	# Maternal heterozygous loci	% Retained maternal heterozygosity in parthenogens	Parthenogenesis mechanism
Shibata et al. (2017)	4	2	0.0%	100%	14	0.0%	Gametic Duplication
Ryder et al. (2021)	2	10 to 11	–	100%	11 to 12	0.0%	Gametic Duplication
Dudgeon et al. (2017)	12	1	0.0%	100%	13	0.0%	Gametic Duplication
Levine et al. (2021)	5	1,445	43.3% to 56.2%	–	–	–	–
Levine et al. (2022)	14	2,074 to 2,227	78.2% to 95.2%	–	–	–	–
Card et al. (2021)	2	20,284	–	99.4% to 99.6%	278	23.0 to 27.0%	Terminal Fusion
Booth et al. (2023)	1	1,86,240	–	99.9%	5,015	3.0%	Terminal Fusion
Levine et al. (2024)	3	4,087	31.1% to 39.2%	–		–	–

All parthenogens displayed homozygosity at >99.4% maternal homozygous loci. Those parthenogens called as the product of gametic duplication retained 0% of maternal heterozygosity, whereas those called as the product of terminal fusion automixis retained 3% to 27% of maternal heterozygosity. All analyses of published data generated results consistent with those previously reported.

as such. In Part 2 of ParthenoGenius, the mechanism of parthenogenesis for all inferred parthenogens was called consistently regardless of whether a user-defined error rate or program-updated error rate was used.

Discussion

We developed *ParthenoGenius*, a Python program that uses heuristics to infer the presence and mechanism of FP from even large genetic/genomic datasets comprising many mothers and their offspring. An attractive feature of *ParthenoGenius* is its ease of use. Only a single input file containing locus names and maternal and offspring genotypes is required, and the only optional arguments that the user may provide are the estimated sequencing error rates for the dataset for Parts 1 and 2. R scripting skills are not required to execute the program. By contrast, other methods for inferring the presence and mechanism of parthenogenesis from genomic parentage data (e.g. bootstrapping estimates of homozygosity and relatedness estimators [Card et al. 2021]), require competency in R scripting and statistics. Moreover, a *ParthenoGenius* file preparation script is provided on GitHub to parse population-wide genetic parentage datasets into suitable input files for *ParthenoGenius*, and a wrapper script is provided to iteratively run *ParthenoGenius* on all files of mother–offspring pairs. The wrapper script takes as command-line arguments the input file directory, output file directory, and user-defined error rates for Part 1 and (optionally) Part 2 of *ParthenoGenius*. The input file format for *ParthenoGenius* is similar to that of the *Structure* input file format (Pritchard et al. 2000), which is a commonly available export format for bioinformatics software (e.g. *Stacks* [Catchen et al. 2011, 2013]) and which may already be in the possession of researchers who have conducted genetic structure analyses. To create the input file required for *ParthenoGenius*, the researcher can simply delete the columns in their structure file that do not contain genotypes or the sample ID. Users who do not have a Structure file but rather a VCF file or other common bioinformatics file format can use PGDspider (Lischer and Excoffier 2011), software that uses a graphical user interface, to convert their file to a Structure file. Finally, *ParthenoGenius* accommodates missing genotypes (coded as –9), such that preanalysis data filtering to remove loci or individuals with missing data is unnecessary (e.g. via *PLINK* [Purcell et al. 2007]).

Since *ParthenoGenius* employs heuristics, all of our analyses ran quickly (Table 2). For example, a scan of 30 *H. charlesbogerti* offspring, each of which was genotyped at over 3,000 loci, was completed by *ParthenoGenius* in approximately 30 s on a personal laptop (2020 MacBook Pro with 2 GHz Quad-Core Intel Core-5 Processor with a solid-state drive).

Limitations and considerations

As a heuristic, *ParthenoGenius* has limitations that do not apply to more complicated and statistically intensive methods of detecting the presence and mechanism of parthenogenesis. First, *ParthenoGenius* requires that the input genetic dataset contains maternal loci that are homozygous, as Part 1 of *ParthenoGenius* relies on a scan of maternal homozygous loci. This requirement is likely to be met in large SNP datasets but may not be in small microsatellite datasets. As a result, small microsatellite datasets in which the mother is heterozygous at all loci cannot be analyzed with *ParthenoGenius*. For example, in our analysis of an *S. fasciatum* set comprising 14 microsatellite loci (Dudgeon et al. 2017), one mother–offspring pair were eliminated from analyses due to a lack of maternal homozygosity.

Additionally, *ParthenoGenius* results are inherently related to the choice of error rate. Overall, our sensitivity tests demonstrated that inferences were robust to the choice of error rates for Parts 1 and 2 of *ParthenoGenius*. However, *O. hannah* parthenogens were erroneously called as nonparthenogens at a very low error rate (0.001), with one called as a nonparthenogen at an error rate of 0.005 as well. These error rates may thus be unrealistically low for this ddRADseq dataset. Indeed, error rates for ddRADseq and RADseq data are dependent on DNA quality, coverage, locus assembly method (i.e. reference-aligned vs. de novo), and qualities of base calls, among other factors. For example, Fountain et al. (2016) used Mendelian errors (i.e. mismatches between parent and offspring genotypes) to estimate error rates for ddRADseq datasets and found that reference-aligned ddRADseq data, such as that of the *O. hannah* study, had error rates ranging from ≤0.01 to ≥0.03 depending on the coverage. The *O. hannah* study (Card et al. 2021) did not

Table 2. We analyzed six published datasets with *ParthenoGenius* to test whether the FP results were congruent with those previously reported.

Species	Parthenogenesis reported	Parthenogenesis supported by ParthenoGenius	Number of offspring	Marker type	Number of loci per sample	Run time with ParthenoGenius wrapper script (s)
Eunectes murinus	Yes	Yes	4	Microsatellite	16	3.29
Gymnogyps californianus	Yes	Yes	2	Microsatellite	22	1.81
Stegostomaz fasciatum	Yes	Yes	12	Microsatellite	14	10.39
Crotalus atrox	No	No	5	ddRAD/SNPs	4,228	5.57
Heloderma charlesbogerti	No	No	14	ddRADseq/SNPs	3,094	30.28
Ophiophagus hannah	Yes	Yes	2	ddRADseq/SNPs	20,562	8.55
Crocodylus acutus	Yes (used PG)	Yes	1	WGS/SNPs	1,91,255	209.23
Ovophis monticola	No (used PG)	No	3	ddRADseq/SNPs	9,697	6.55

We also analyzed two novel datasets (Booth et al. 2023; Levine et al. 2024) to test for FP. Run times reflect the time that the *ParthenoGenius* wrapper script took to analyze all mother–offspring datasets locally on a personal laptop (2020 MacBook Pro with 2 GHz Quad-Core Intel Core-5 Processor with a solid-state drive).

report the coverage of the ddRADseq loci, so we are unable to evaluate the effect of coverage on the error rate for these data.

Importantly, though, all previously reported nonparthenogens were called as such regardless of the error rates tested. This is because the distinction between nonparthenogens and parthenogens with regard to the presence of alternate offspring alleles at maternal homozygous loci is rather large. For example, among the SNP data analyzed for nonparthenogens, the mean proportion of offspring genotypes that differed from maternal genotypes at maternal homozygous loci ranged from 0.169 in *H. charlesbogerti* to 0.519 in *C. atrox*. By comparison, the proportion of offspring genotypes that differed from the maternal genotypes at maternal homozygous loci in *C. acutus* was 0.00045. Similarly, the mean proportion of offspring genotypes at maternal homozygous loci that differed from the maternal genotype in *O. hannah* parthenogens was 0.0048.

An additional consideration is that *ParthenoGenius* may struggle to accurately call the mechanism of FP if the percent retained maternal heterozygosity for the offspring approximates the cutoff value between terminal fusion and central fusion automixis (i.e. 67%). However, it is quite unlikely that terminal fusion would result in such high retained heterozygosity at maternal loci. The reasoning behind this is that at the centromere, recombination will be close to zero, increasing to a theoretical maximum of 67% toward only the terminal tips (Svendsen et al. 2015). As such, recent studies investigating FP using genome-wide markers report low levels of parthenogen-observed genome-wide heterozygosity (e.g. Card et al. 2021- *O. hannah*—0.007; Booth et al. 2023—*C. acutus*—0.045). Therefore, differentiating between terminal fusion and central fusion should not prove to be an issue. Nonetheless, the user should approach interpretation of their results with caution if the percentage of retained maternal heterozygosity in offspring nears this cutoff value.

Users should note that *ParthenoGenius* does not provide an assessment of confidence in calls of FP or FP mechanism. This is by design to allow the user to quickly screen large parentage datasets for the presence of candidate parthenogens. If parthenogens are called by *ParthenoGenius* and the user desires additional support for the inferences, we suggest that the user can then further evaluate the genomes of the inferred parthenogens using more rigorous and traditional statistical methods (e.g. *Rhh* [Alho et al. 2010]; *Demerelate* [Kraemer and Gerlach 2017]). However, we emphasize that *ParthenoGenius* called the presence and mechanism of parthenogenesis consistent with previously reported results using a simple and intuitive method in all of the published datasets analyzed. Furthermore, parthenogenesis called in the novel *C. acutus* dataset (Booth et al. 2023) was supported by corroborating data that implicated FP (i.e. long-term isolation of the mother and high clutch failure).

Discriminatory power of microsatellites to detect FP

Importantly, one should consider the discriminatory power of their dataset to infer FP and its mechanisms. In small genetic datasets such as microsatellite data, few maternal loci may be homozygous (e.g. one or two), and nonparthenogens may be homozygous at the few maternal homozygous loci sequenced if both parents had the same alleles at those loci (i.e. not due to parthenogenesis but rather a function of population allele frequencies). In this sense, the small number of maternal homozygous loci in the data results in low power to detect FP with *ParthenoGenius*, and this could cause *ParthenoGenius* to erroneously call nonparthenogens as parthenogens. We note, however, that the heterozygous loci in microsatellite datasets may have adequate power to detect parthenogens via other methods.

Similarly, *ParthenoGenius* may call the mechanism of parthenogenesis as gametic duplication in small datasets simply because loci that display retained maternal heterozygosity were not genotyped. This is likely in small microsatellite datasets as retained heterozygosity in parthenogens produced via terminal fusion automixis may be restricted to only the tips of chromosomes (Lampert 2008). For example, all of the parthenogens in the microsatellite datasets that we analyzed (i.e. Dudgeon et al. 2017; Shibata et al. 2017; Ryder et al. 2021) were called as the product of gametic duplication by *ParthenoGenius*. While gametic duplication may be the mechanism of FP that produced the parthenogens in these studies, it is also likely that these small microsatellite datasets (i.e. 14 to 22 loci) lack the power to detect other mechanisms. The user should, therefore, carefully consider the results of *ParthenoGenius* in the context of the discriminatory power of their data, with genome-wide datasets more likely to result in accurate mechanistic inference.

Conclusions

Due to its intuitive logic, ease of use, and ability to screen large genomic parentage datasets quickly, *ParthenoGenius* will facilitate rapid screening of genetic/genomic parentage datasets for the presence and mechanism of FP. We suggest that *ParthenoGenius* could be incorporated by the user into their parentage analysis workflows or used to analyze existing genomic parentage data to screen for parthenogens in taxa for which FP is likely to occur (e.g. crocodilians, reptiles, birds, and elasmobranch fishes). In this sense, *ParthenoGenius* promises to improve our understanding of the frequency of FP and LTSS in captive and wild populations and may prove to be an important tool in the development of conservation and management strategies.

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Funding

This research was supported by a faculty startup from Kean University to BAL and Division of Environmental Biology 1754394 to WB. Additional funding came from faculty startup from Virginia Polytechnic Institute and State University to WB.

Conflict of interest statement. None declared.

Author contributions

Brenna Levine (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing—original draft, Writing—review & editing) and

Warren Booth (Conceptualization, Funding acquisition, Methodology, Supervision, Writing—review & editing)

Data availability

The ParthenoGenius.py script, a wrapper script that iteratively runs ParthenoGenius on multiple mother/offspring pairs (PG_wrapper.sh), a data file preparation script that generate ParthenoGenius input files from large population genetic files (PG_file_prep.sh), all input and output data files, and ParthenoGenius documentation are publicly available on GitHub (https://github.com/brenna-levine/LEVINE_ParthenoGenius).

References

Alho JS, Valimaki K, Merlia J. Rhh: an R extension for estimating multilocus heterozygosity and heterozygosity-heterozygosity correlation. Mol Ecol Res. 2010:10:720–722.

Allen L, Sanders KL, Thomson VA. Molecular evidence for the first records of facultative parthenogenesis in elapid snakes. Open Sci. 2018:5:171901.

Andrews KR, Adams JR, Cassirer EF, Plowright RK, Gardner C, Dwire M, Hohenloe PA, Waits LP. A bioinformatic pipeline for identifying informative SNPs panels for parentage assignment from RADseq data. Mol Ecol Res. 2018:18:1263–1281.

Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman J, Li Z, St. Leger J, Shao C, et al. Evolution of germline mutation rates among vertebrates. Nature. 2023:615:285–291.

Booth W, Johnson DH, Moore S, Schal C, Vargo EL. Evidence of viable, non-clonal but fatherless Boa constrictors. Biol Lett. 2011a:7:253–256.

Booth W, Levine BA, Corush JB, Davis MA, Dwyer Q, De Plecker R, Schuett GW. Discovery of facultative parthenogenesis in a new world crocodile. Biol Lett. 2023:19:20230129.

Booth W, Million L, Reynolds RG, Burghardt GM, Vargo EL, Schal C, Tzika AC, Schuett GW. Consecutive virgin births in the new world boid snake, the Colombian rainbow boa, *Eipcrates maurus*. J Hered. 2011b:102:759–763.

Booth W, Schuett GW. Molecular genetic evidence for alternate reproductive strategies in North American pitvipers (Serpentes: Viperidae): long-term sperm storage and facultative parthenogenesis. Biol J Linn Soc. 2011:104:934–942.

Booth W, Schuett GW. The emerging phylogenetic pattern of parthenogenesis in snakes. Biol J Linn Soc. 2016:118:172–186.

Booth W, Schuett GW, Ridgway A, Buxton DW, Castoe TA, Bastone G, Bennett C, McMahan W. New insights on facultative parthenogenesis in pythons. Biol J Linn Soc. 2014:112:461–468.

Booth W, Smith CF, Eskridge PH, Hoss SK, Mendelson III JR, Schuett GW. Facultative parthenogenesis discovered in wild vertebrates. Biol Lett. 2012:8:983 –9985.

Card DC, Vonk Freek J, Smalbrugge S, Casewell NR, Wuster W, Castoe TA, Schuett GW, Booth W. Genome-wide data implicate terminal fusion automixis in king cobra facultative parthenogenesis. Sci Rep. 2021:11:7271.

Catchen J, Amores A, Hohenloe P, Cresko W, Postlethwait J. Stacks: building and genotyping loci *de novo* from short-read sequences. G3: Genes Genom Genet. 2011:1:171–182.

Catchen J, Hohenloe P, Bassham S, Amores A, Cresko W. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013:22:3124–3140.

Chapman DD, Firchau B, Shivji MS. Parthenogenesis in a large-bodied requiem shark, the blacktip *Carcharhinus limbatus*. J Fish Biol. 2008:73:1473–1477.

Chapman DD, Shivji MS, Sommer J, Fletcher H, Prodöhl PA. Virgin birth in a hammerhead shark. Biol Lett. 2007:3:425–427.

Dudgeon CL, Coulton L, Bone R, Ovenden JR, Severine T. Switch from sexual to parthenogenetic reproduction in a zebra shark. Sci Rep. 2017:7:40537.

Feldheim KA, Chapman DD, Sweet D, Fitzpatrick S, Prodöhl PA, Shivji MS, Snowden B. Shark virgin birth produces multiple, viable offspring. J Hered. 2010:101:374–377.

Fields AT, Feldheim KA, Poulakis GR, Chapman DD. Facultative parthenogenesis in a critically endangered wild vertebrate. Curr Biol. 2015:25:R446–R447.

Fountain ED, Pauli JN, Reid BN, Palsboll PJ, Peery Z. Finding the right coverage: the impact of coverage and sequencing quality on single nucleotide polymorphism error rates. Mol Ecol Res. 2016:16:966–978.

Germano DJ, Smith PT. Molecular evidence for parthenogenesis in the Sierra garter snake, *Thamnophis couchii* (Colubridae). Southwest Nat. 2010:55:280–282.

Harmon TS, Kamerman TY, Corwin AL, Sellas AB. Consecutive parthenogenetic births in a spotted eagle ray *Aetobatus narinari*. J Fish Biol. 2016:88:741–745.

Holt WV, Loyd RE. Sperm storage in the vertebrate female reproductive tract: how does it work so well? Theriogenology. 2010:73:713–722.

Huisman J. Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. Mol Ecol Resour. 2017:17:1009–1024.

Jones O, Wang J. COLONY: a program for parentage and sibship inference from multilocus genotype data. Mol Ecol Resour. 2010:10:551–555.

Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program Cervus accommodates genotyping error increases success in paternity assignment. Mol Ecol. 2007:16:1099–1106.

Kess T, Gross J, Harper F, Boulding EG. Low-cost ddRAD method of SNP discovery and genotyping applied to the periwinkle *Littorina saxatilis*. J Molluscan Stud. 2016:82:104–109.

Kinney ME, Wack RF, Grahn RA, Lyons L. Parthenogenesis in a Brazilian Rainbow Boa (Epicrates cenchria cenchria). Zoo Biology. 2012:32:172–176. doi:10.1002/zoo.21050

Kono T. Genomic imprinting is a barrier to parthenogenesis in mammals. Cytogenet Genome Res. 2006:113:31–35.

Kraemer P, Gerlach G. *Demerelate*: calculating interindividual relatedness for kinship analysis based on codominant diploid genetic markers using R. Mol Ecol Resour. 2017:17:1371–1377.

Kratochvíl L, Vulkić J, Červenka J, Kubička L, Pokorná MJ, Kukačková D, Rovatsos M, Piálek L. Mixed-sex offspring produced via cryptic parthenogenesis in a lizard. Mol Ecol. 2020:29:4118–4127.

Lampert KP. Facultative parthenogenesis in vertebrates: reproductive error or chance? Sex Dev. 2008:2:290–301.

Levine BA, Douglas MR, Yackel Adams A, Lardner B, Reed RN, Savidge JA, Douglas ME. Genomic pedigree reconstruction identifies predictors of mating and reproductive success in an invasive vertebrate. Ecol Evol. 2019:9:11863–11877.

Levine BA, Hill RL, Mendelson III JR, Booth W. Parentage assignment reveals multiple paternity in the critically-endangered Guatemalan beaded lizard (*Heloderma charlesbogerti*). Cons Gen. 2022:23:859–863.

Levine BA, Moresco A, Trout T, Schuett GW, Booth W. Female long-term sperm storage results in viable offspring in the Himalayan Mountain Pitviper, *Ovophis monticola*. Zoo Biol. 2024:43:183–187. doi:10.1002/zoo.21820

Levine BA, Schuett GW, Booth W. Exceptional long-term sperm storage by a female vertebrate. PLoS One. 2021:16:e0252049.

Lischer HEL, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2011:28:298–299. doi:10.1093/bioinformatics/btr642

Miller KL, Rico SC, Muletz-Wolz CR, Campana MG, McInerney N, Augustine L, Frere C, Peters AM, Fleischer RC. Parthenogenesis in a captive Asian water dragon (*Physignathus cocincinus*) identified with novel microsatellites. PLoS One. 2019:14:e0217489.

Olson MW. Nine year summary of parthenogenesis in turkeys. Proc Soc Exp Biol Med. 1960:105:279–281.

Parker HM, McDaniel CD. Parthenogenesis in unfertilized eggs of *Coturnix chinensis*, the Chinese painted quail, and the effect of egg clutch position on embryonic development. Poult Sci. 2009:88:784–790.

Pearcy M, Hardy OJ, Aron S. Thelytokous parthenogenesis and its consequences on inbreeding in an ant. Heredity. 2006:96:377–382.

Pearcy M, Hardy OJ, Aron S. Automictic parthenogenesis and rate of transition to homozygosity. Heredity. 2011:107:187–188.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One. 2012:7:e37135.

Portnoy DS, Hollenbeck CM, Johnston JS, Casman HM, Gold JR. Parthenogenesis in a whitetip reef shark *Triaenodon obesus* involves a reduction in ploidy. J Fish Biol. 2014:85:502–508.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000:155:945–959.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007:81:559–575.

Ramachandran R, McDaniel CD. Parthenogenesis in birds: a review. Reproduction. 2018:155:R245–R257.

Reynolds GR, Booth W, Schuett GW, Fitzpatrick BM, Burghardt GM. Successive virgin births of viable male progeny in the checkered gartersnake, *Thamnophis marcianus*. Biol J Linn Soc. 2012:107:566–572.

Robinson DP, Baverstock W, Al-Jaru A, Hyland K, Khazanehdari KA. Annually recurring parthenogenesis in a zebra shark *Stegostoma fasciatum*. J Fish Biol. 2011:79:1376–1382.

Ryder OA, Thomas S, Judson JM, Romanov MN, Dandekar S, Papp JC, Sidak-Loftis LC, Walker K, Stalis IH, Mace M, et al. Facultative parthenogenesis in California condors. J Hered. 2021:112:569–574.

Saunders IW, Brohede J, Hannan GN. Estimating genotyping error rates from Mendelian errors in SNP array genotypes and their impact on inferences. Genomics. 2007:90:291–296.

Schuett GW, Fernandez PJ, Gergits WF, Casna NJ, Chizar D, Smith HM, Mitton JB, Mackessy SP, Odum RA, Demlong MJ. Production of offspring in the absence of males: evidence for facultative parthenogenesis in bisexual snakes. Herpetol Nat Hist. 1997:5:1–10.

Schut E, Hemmings N, Birkhead TR. Parthenogenesis in a passerine bird, the zebra finch *Taeniopygia guttata*. Ibis. 2008:150:197–199.

Seixas F, Morinha F, Luis C, Alvura N, Pires MDA. DNA-validated parthenogenesis: first case in a captive female Cuban boa (*Chilobothrus angulifer*). Salamandra. 2020:56:83–86.

Shibata H, Sakata S, Hirano Y, Nitasaka E, Sakabe A. Facultative parthenogenesis validated by DNA analyses in the green anaconda (*Eunectes murinus*). PLoS One. 2017:12:e0189654.

Solter D. Differential imprinting and expression of maternal and paternal genomes. Annu Rev Genet. 1988:22:127–146.

Straube N, Lampert KP, Geiger MF, Weiß JD, Kirchhauser JX. First record of second-generation facultative parthenogenesis in a vertebrate species, the whitespotted bambooshark *Chiloscyllium plagiosum*. J Fish Biol. 2016:88:668–675.

Svendsen N, Reisser CMO, Dukić M, Thuillier V, Ségard A, Liautard-Haag C, Fasel D, Hürlimann E, Lenormand T, Galimov Y, et al. Uncovering cryptic asexuality in *Daphnia magma* by RAD sequencing. Genetics. 2015:201:1143–1155.

Thrasher DJ, Butcher BG, Campagna L, Webster MS, Lovette IJ. Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: a proof of concept in a highly promiscuous bird. Mol Ecol Res. 2018:18:953–965.

Watts PC, Buley KR, Sanderson S, Boardman W, Ciofi C, Gibson R. Parthenogenesis in Komodo dragons. Nature. 2006:444:1021–1022.