Dependence in genome descent in small populations

Elizabeth Thompson*
University of Washington, Seattle, USA- eathomp@uw.edu

Abstract

This paper presents a small part of an ongoing study to understand the processes and probabilities of genome descent and ancestry within the context of the population pedigrees of small populations. Here we focus on genome survival and descent from founders, using as an example a random-mating population of a diploid dioecious species of 100 males and 100 females in each generation. We consider the descendants of individuals, and then of their genomes, and specifically of the segments of founder genomes represented in the population after varying numbers of generations. We consider the strong positive dependence in survival and descent between the two genomes of founders, and the way in which initial mating structure may impact the dependence between the genomes of founder mates. We discuss these last results in the context of two species, rescued from very small remnant populations: the Przewalski horse and the California condor.

Keywords: genome descent; small populations; dependence in survival; endangered species;

1. Introduction

This paper presents a small part of an ongoing study to understand the processes and probabilities of genome descent and ancestry in small populations, with the goal of understanding how genetic diversity may best be maintained in highly endangered species. Both ancestry and descent are of interest. In the species the current individuals may be remnants of large populations, with only a small part of that broad ancestral diversity remaining. Conversely the remnant population will be the founders of the species for the future. Descent from them will define what diversity can be retained.

Under the process of meiosis, segments of genome are copied from generation to generation. The genome of a current individual consists of segments of ancestral DNA from ancestors at a specified time-depth, or from founders of a defined pedigree. While classical population genetics provides many results relating to single genome locations, such as measures of inbreeding and heterozygosity, the effect of genome length, chromosomal structure, and the segmental nature of inheritance has been much less studied.

While the initial abstract for WSC-2025 was very broad, mentioning aspects of ancestry and descent across genomes, across geographical space in structured populations, and over time, only a small part can be addressed in this short contribution. The case of ancestry of a genome of a current individual was fully considered in Thompson (2024). Here we develop further the results of Thompson(2023), considering the descent and survival of founder genomes in a small population. Specifically, for the purposes of presentation of results, we will consider a diploid dioecious species, in a random-mating population of 100 males and 100 females each generation. We will use an idealized genome of genetic length in total 30 Morgans, and, where necessary, divide this genome into 30 chromosomes each of 1 Morgan length. In broad terms this provides a good approximation for many mammals. We emphasize there are no genetic data here; the goal is understanding the processes of ancestry and descent and of the loss of genetic variation.

2. Individual diploid extinction and descent

We consider first extinction and survival of the descendants of a diploid individual. In an unbounded population, in which each individual has a Poisson number of offspring, with mean 2 ($\mathcal{P}o(2)$), simple branching process computations provide that the probability of extinction rises within 5 generations to just over 20%, and thereafter remains effectively constant. In contrast, the number of copies of a single genome location (or "gene") at the next generations is $\mathcal{P}o(1)$. The probability of extinction converges eventually to 1, but very slowly; the expected time to extinction is infinite (Figure 1).

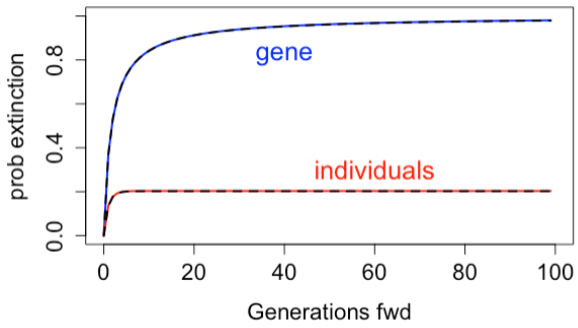

Figure 1: Extinction of the descent of a diploid individual, and of DNA at a single genome location

These extinction probabilities are almost unaffected by population size. Also shown on Figure 1 are black dashed lines showing the values for a random-mating population of 100 males and 100 females; these are almost indistinguishable from the Poisson case. Of course, at the very longest times and in the very smallest populations, extinction probabilities are smaller: some descendant lineages must survive in an extant population.

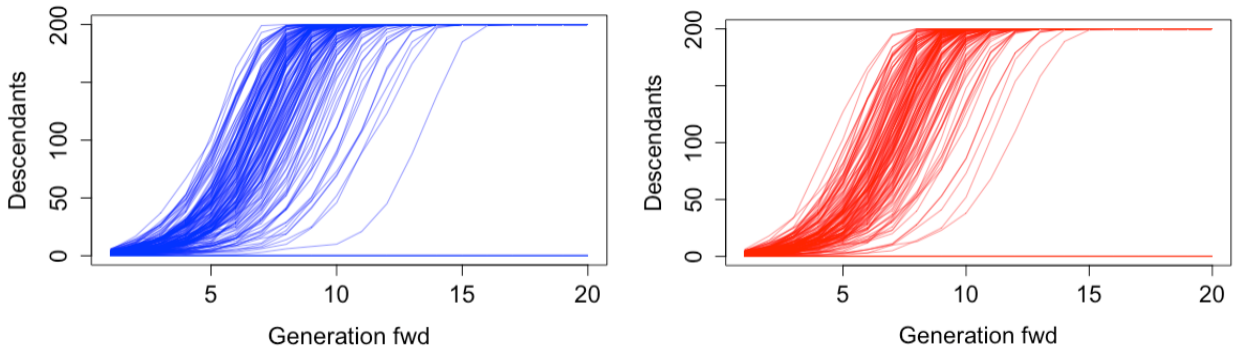

Figure 2: Descendant counts in a random–mating population of 100 males and 100 females. Left: 200 independent realization. Right: 200 individuals of a single population realization.

Figure 2 shows the descendant counts from founder individuals in a random mating population of 100 males and 100 females. On the left is shown 200 independent realizations of the process whereas on the right are the values for each of the 200 founders in a single population. There is more variation in descent than ancestry (Thompson 2013, 2024), but the time frame is very similar. Within 17 generations each founder is either an ancestor of all 200 individuals, or of none. Further, extinction is very rapidly determined: individuals with descendants at 5 generations will almost certainly become ancestors of everyone, although in a few cases descendant counts remain low for more generations.

Figure 3 shows the standard deviations (SD) of the descendant counts for the between and within-population realizations of Figure 2, including only the individuals with a positive number of descendants. It is seen that while the within-population SD is smaller, there is little difference. Overall, there seems little dependence between founders in either their survival or in the number of their descendants.

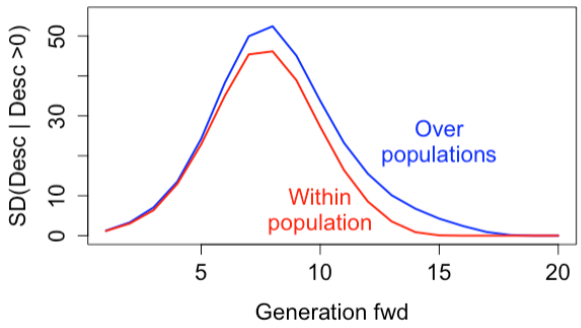

Figure 3: Standard deviations of descendant counts: see text for details..

3. The survival and descent of genome

To consider the descent of genome we first introduce the process of meiosis (Figure 4). For each autosome, the parent has two copies, one paternal (blue), one maternal (red). The pairs of double-stranded DNA chromosomes of the parent duplicate and align forming the tetrad (8 strands). At this stage chiasmata (designated by black crosses) may form, permitting *crossovers* of DNA between maternal and paternal homologues.

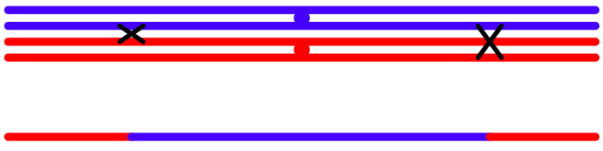

Figure 4: An offspring chromosome resulting from a parental meiosis

There are four potential offspring chromosomes, one of which is shown. In this example, the offspring chromosome results from two crossovers and consists of a segment of maternal DNA, then paternal, and then maternal. Thus chromosomes are inherited in large segments. In units of genetic distance, a good model is that crossovers occur as a Poisson process, rate 1 per Morgan. The relationship between genetic distance and base-pair (bp) distance is complex and variable, but overall 1 Morgan corresponds to about 10^8 bp. The Poisson process model is good for all but the smallest chromosomes. For meiosis to function, there should be at least one chiasma on the tetrad, so every offspring chromosome contains a crossover with probability at least 0.5, and hence has a genetic length of at least 0.5 Morgans. However, in small chromosomes, the probability of more than one crossover is less than given by the Poisson model. As founder chromosome segments are broken up and dispersed in the population, this departure from the model has decreasing effect.
For N individuals ($2N$ founder genomes), each of length L Morgans, subdivided into C chromosomes, after k generations the total number of segments of founder genome in the population is Poisson $\mathcal{P}o(2N(C + kL))$, or $\mathcal{P}o(2N((k + 1)L)$ if $C = L$. To a remarkably close approximation each had an exponential length, mean $1/(k + 1)$ (Thompson, 2024), and the overall probability a segment is broken at the next generation is

$$\int_0^\infty (k+1)\ \exp(-(k+1)x)\ (1-\exp(-x))\ dx \quad = \quad 1 \ - \ \frac{k+1}{k+2} \quad = \quad \frac{1}{k+2}$$

Figure 5 shows the probability of extinction of a segment of DNA of a given genetic length. Because of the segmental inheritance of DNA, a segment of 0.1 Morgan (of order 10^7bp) has high probability of extinction. In the first few generations it behaves almost like a single point in the genome. Only as it gets broken up and dispersed does the probability of survival of some part of it improve. At the other extreme, for a 30 Morgan genome the probability that some part will survive is close to the probability of having descendants over 40 generations.

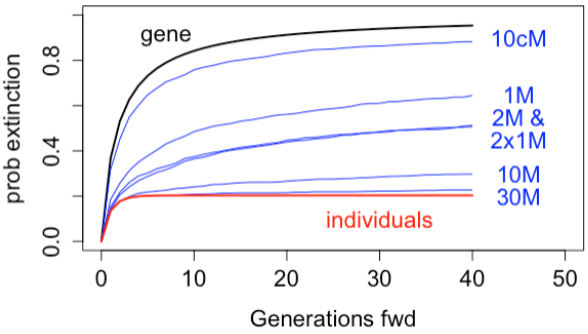

Figure 5: Probabilities of extinction of the entirety of a given length of genome.

However it is possible to have descendants but no genetic descendants, and the extinction probability for a 30M genome continues to increase over time. A genome length of 1 to 2 Morgans (typical of a mammalian chromosome) has an extinction probability around 50% over 40 generations, but these extinction probabilities also continue to increase over time. Also shown in Figure 5 are two lines, one for a segment length 2 Morgans, and the other for two 1-Morgan lengths. The latter has very slightly lower extinction probability in the initial generations, but overall the exact division of genome into chromosomes has negligible effect.

4. Descendant segments and descendant genomes

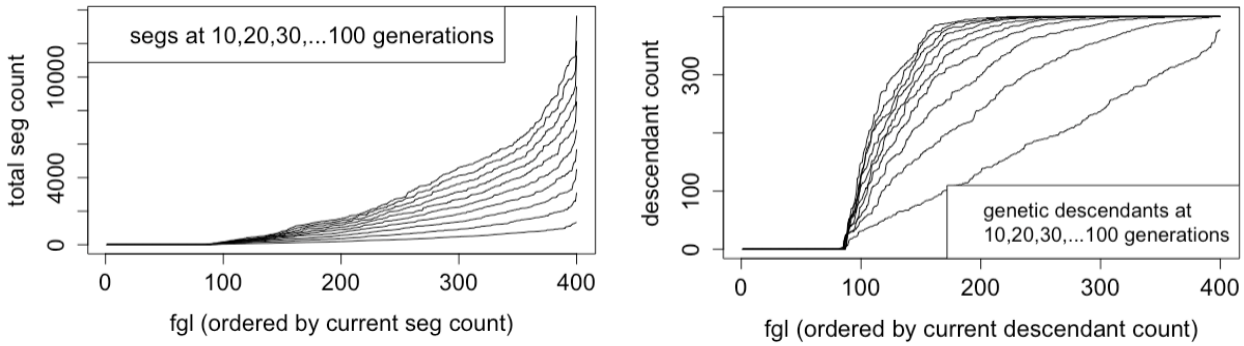

Figure 6: Descendant counts of the 400 founder genomes Left: counts of descendant segments. Right: Counts of current genomes that each founder genome is represented in.

We consider now the way in which the $\mathcal{P}o(2N(k+1)L)$ total descendants genome segments at generation k are attributable to each of the $2N = 400$ founder genomes, and how many of the potential $2N = 400$ current genomes each of the 400 founder genomes is represented in. Figure 6 shows these counts in increasing order at intervals of 10 generations from 10 to 100. First, just over 20% of founder genomes rapidly become extinct. They have no current segments (left) and are present in no current genomes (right). Initially (at 10 generations) the number of descendant segments is fairly uniform among founder genomes with genetic descendants (left) as is the count of the genetic descendant genomes (right). Over time, there become a few genomes with a very large number of descendant segments. Of course, many of these are segments covering the same

portions of the genome. It is usually not the case that a large portion of the founder genome survives: that aspect is not considered in this paper. At 100 generations the median number of descendant segments is about 2000, but some founder genomes have few surviving descendant segments. However, by 100 generations the 50% of founder genomes with above median segment counts are typically present in all 400 current genomes. Again, there are also non-extinct founder genomes with segments in only a few current individuals.

Figure 7 gives a slightly different view of the same results, with descendants scored by individual rather than by haploid genome. For a single random-mating population of 100 males and 100 females, in red are shown the number of descendants over the generations. (These are the same red curves as in Figure 2 extended now to 40 generations.) In blue are the numbers of genetic descendants – individuals carrying DNA from each founder. Counts of genetic descendants rise more slowly than of descendants. Although many founders are genetic ancestors of all 200 individuals by generation 15, there are about 10 individuals who are ancestors of everyone, but genetic ancestors of few.

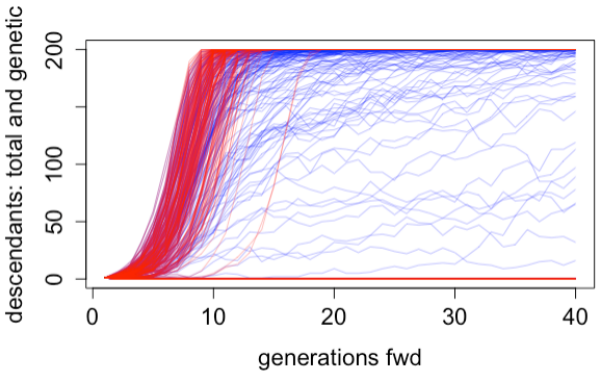

Figure 7: Comparison of the number of descendants vs number of genetic descendants.

5. Dependence in founder genome descent

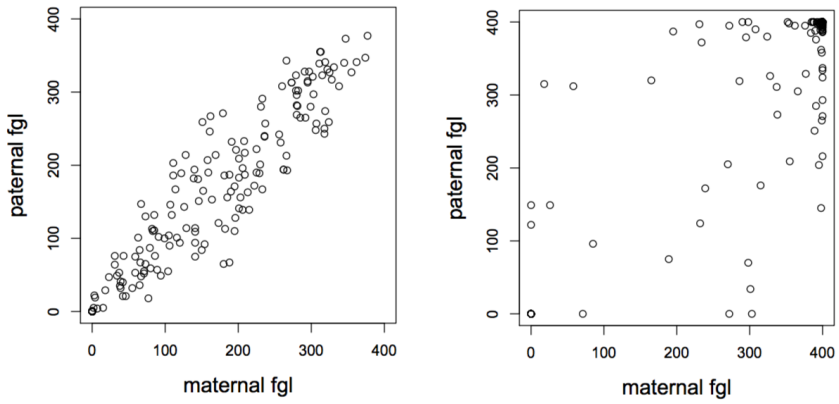

Figure 8: Number of descendant genomes of each founder's maternal and paternal genomes in a random-mating population of 100 males and 100 females. Left: Generation 10. Right: Generation 100.

While overall the dependence in descent among founders is slight (Figure 3) some dependencies are strong. One such is the dependence between the two haploid genomes of each diploid founder. Figure 8 shows scatterplots of the number of haploid genomes at generation 10 and at generation 100 that are genetic descendants of maternal and paternal founder genomes (fgl). The correlation is

not only very high (0.9 to 0.85) but is maintained over the generations. This results from multiple junctions formed between the two genomes at the very first generation– on average $L = 30$ per offspring. These junctions descend as single location markers, carrying the two genomes together. While many are lost, a few survive, and become frequent in the descendants.

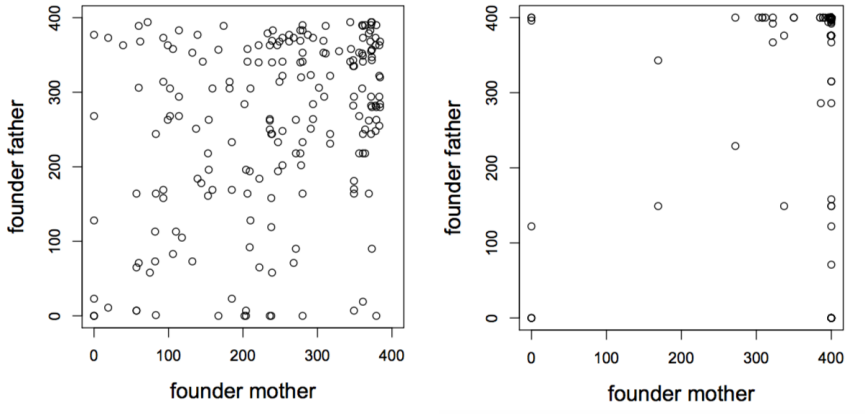

Figure 9: Number of descendant genomes of initial-generation mates in a random-mating population of 100 males and 100 females. Left: Generation 10. Right: Generation 100.

The question then arises: what about the next generation? Figure 9 gives a similar scatterplot, this time for the counts of genomes that are genetic descendants of either genome of founder mates. A positive correlation exists but is far lower (0.25 to 0.28). However it is again stable over time. This correlation results from junctions between the founder-mate genomes formed in meioses from their offspring. However, in a random-mating population an individual with many offspring will have many different mates, who may have few offspring, or many but with other mates.

Offspring assignment to to parent couples	Number of descendant genomes	Number of descendant segments
$2N$ random couples; with replacement Each offspring randomly assigned	0.25 to 0.28	0.16 to 0.29
N random couples ; with replacement 1 son, 1 daughter each	0.33 to 0.35	0.22 to 0.24
N couples sampled; no replacement $2N$ offspring independently assigned to a couple	0.85 to 0.95	0.83 to 0.92

Table 1: Correlations in descendant counts of genomes from founder mates in populations of 100 males and 100 females

Table 1 shows the correlations in descent from founder mates under three different forms of random mating in the population of 100 males and 100 females at each generation. Shown are both correlations in the number of descendant genomes in which either genome of each founder genome is represented, and the total number of genome segments descending from either genome of each founder. The numbers shown are approximate from a small number of realizations. They are

broadly constant over time, since they derive from junctions formed in meioses of the immediate offspring of founders. The form of mating does not impact correlations between the two genomes of each founder: these remain at 0.85 to 0.95 for the count of descendant genomes, and 0.8 to 0.9 for descendant genome segments.

The first line of the table corresponds to true random mating in a dioecious population (Figure 9). Each of $2N$ offspring is assigned a random father and a random mother, each sampled with replacement from the parental males and females. In the second case, each of N male offspring is assigned a random father and a random mother, each sampled with replacement for the parental males and females, and a female offspring is then assigned to the same parent couple. This creates pairs of full siblings with larger sets of half-siblings, but the effect on dependence in descent is only about a 25% increase. In contrast is the bottom row of the table. Here the N couples are randomly formed, but then fixed. Each of $2N$ offspring is independently assigned to a couple. In this case, all offspring fall into full sibships. If a founder has many offspring so does his mate, and at the next generation many junctions will be formed between the genome of the founder mates. Correlations as high as the initial within-founder correlations are restored.

6. Management of endangered species

There are several implications for the management of remnant species when restoration is planned. First overall time scales are short. For a diploid population of 200, within 20 generations, a founder individual is an ancestor either of everyone or of no one. Genetic segment descent is a bit slower, but at 30 generations almost all descendants are genetic descendants, and by 80 generations, 50% of founder genomes are in all 400 current ones.

Second the very first generations are important, ensuring survival of genome from as many individuals as possible, and not allowing any to dominate. As has long been recognized in captive management of species, those with multiple offspring will have increasing dominance over subsequent generations. Loss of genome will always occur, but genome is particularly at risk in the early generations, when founder genome segments are still large, so that large segments can be lost.

Finally the choice of mating pairs can be important, in saving a diversity of genomes. One may contrast the case of the Przewalski horse (Geyer et al, 1989, Thompson 2023) with that of the California condor (Geyer et al., 1993). Both current populations descend from about 12 captive individuals They now number over 2000 and over 500 respectively, with the majority of individuals in reintroduced semi-wild populations. But their genetic histories are very different. Figure 10 shows the ancestry of the animal that was the origin of the first Przewalski horse clone born in August 2020. The source animal has many descendants in the European captive population, but the seven founders from whom he is descended are less well represented in the North American population. More generally, the population remained very subdivided and small until after 1950, leading to much loss of founder genome (Geyer et al., 1989). Moreover there has been introgression from Mongolian domestic horses. Founder #229 was a Mongolian domestic horse, and #18 is suspected to have been a first-generation hybrid.

Of interest here, however, is the question of dependence among founder genomes. While the clone source #615 represents 7 founders, descent is though just four individuals 4 generations later. Although the mates #39,#40 have other offspring descendants in the population, the mates #17,#18 have only the single offspring shown (Volf, 1961 et seq.). Thus wherever a segment from #17 descends to #615, or to the population in general, it will almost always be accompanied by a segment of the hybrid #18. The species contributions of the trio #229,#11,#12 are through their two offspring shown. So again descent of segments from #11, and to a lesser extent from #12, is almost always accompanied by a segment of genome from the Mongolian domestic horse #229.

While we cannot blame the owners of the Przewalski horses of 100 years ago, and must be very thankful that they saved the species from extinction, with hindsight we must admit the genetic management up to 1970 was non-existent. From about 1970, the picture is very much improved. The population increased, a global species survival plan was initiated, and horses were transferred among populations. Indeed, #615 himself, born at Whipsnade (UK) to a Whipsnade dam and a sire transferred there from the Prague Zoo, was moved from the UK to USA in 1978 at age 3, and then within the USA in 1990. Since 1970 there has been little genome loss from the population.

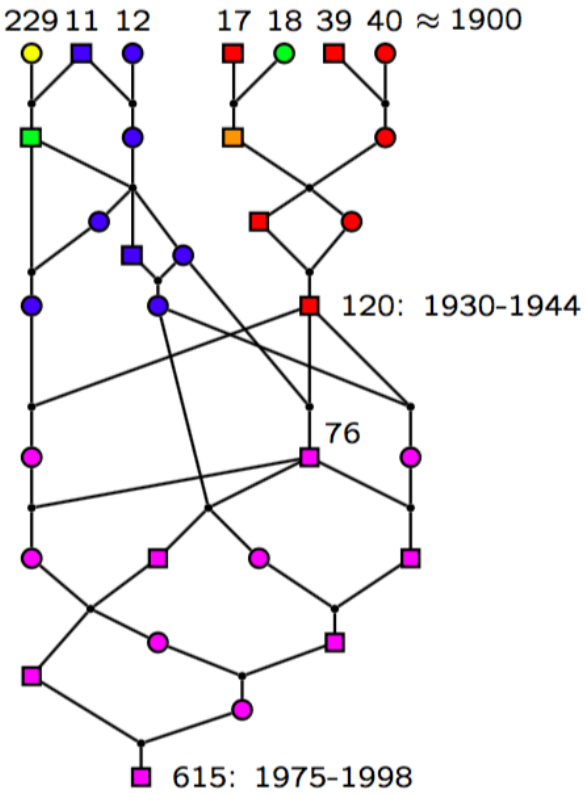

Figure 10: The pedigree of the Przewalski horse clone.

In terms of founder genome survival, the story is much more positive for the California condor, also down to about 12 birds in 1987 after the last adult was brought into captive management from the wild (US Fish & Wildlife Service, 1987). While it took a few years for successful captive breeding to be established, great attention was paid to the establishment of successful breeding pairs. Genetic analysis provided information on the structure of relatedness and diversity in the remnant population (Geyer et al, 1993), and these also informed some of the management decisions made. Here the dependence in survival of the genomes of founder mates will have had a very positive impact on overall diversity. While those early decisions were undoubtedly the most important in maintaining genomic diversity, similar considerations informed the release program when the wild population was reestablished, to ensure both the free-flying birds, and the captive-breeding program could represent surviving genetic diversity.

7. Conclusion

Study of the ancestry and descent of genomes in small populations continues to inform the genetic structures of current populations, and the potential for future maintenance of genetic diversity in severely endangered species, Consideration of the dependence in the survival of the two genomes of diploid founders, and of the genomes of funder mates, can inform the rescue management of remnant populations. However, meiosis is a highly variable process, so that close management of a population pedigree may have limited impact on the pattern of surviving genomes.

Acknowledgment

I am grateful to Dr Oliver Ryder of the San Diego Zoo Wildlife Alliance for introducing me to the pedigrees and studies both of the Przewalski horse and of the California condor, for collaboration over many years, and for his continuing interest in this work. I acknowledge also the important work and tireless energy of Dr. Waltraut Zimmermann of the Cologne Zoo (Germany) in maintaining from 1986 to 2012 the fullest possible records of the complete Przewalski Horse population from its founding to the current day.

References

Geyer C. J., Thompson E. A. and Ryder O. A. (1989) Gene survival in the Asian Wild Horse(Equus Przewalskii): II. Gene survival in the whole population, in subgroups, and through history. Zoo Biology, 8: 313-329.

Geyer, C. J., Ryder, O. A., Chemnick, L. G. and Thompson, E. A. (1993) Analysis of relatedness in the California condors from DNA fingerprints. Mol. Biol. Evol., 10: 571-589.

Thompson, E. A. (2013) Identity by descent: Variation in meiosis, across genomes, and in populations. Genetics 194: 301-326.

Thompson , E. A. (2023) Dependence in the survival of ancestral genomes. Oral Contributed Proceedings of the 64th World Statistics Congress of the International Statistical Institute. Ottawa, Canada.

Thompson, E. A. (2024) The ancestry of a genome. Or, The amazing ubiquity of exponential distributions. .bioRxiv: page 10.1101/2024.08.30.610585, August 2024.

U.S. Fish & Wildlife Service (1987) Last Wild California Condor Capture for Breeding Program. Press release: archived pdf. https://www.fws.gov/news/Historic/NewsReleases/1987/19870421.pdf

Volf J (1961 et seq.) International Studbook of the Przewalski's Horse. Tech. rep., Prague Zoo, Prague.