The curious case of a hapless mathematical contribution to biology

Eric Tannier

A multidisciplinary manuscript.

The curious case of a hapless mathematical contribution to biology

Chromosome inversions in *Drosophila*, 1937-1941

Eric Tannier

Abstract This is the story, revisited by current eyes and means, of a mathematical biology problem explored in the 1930s in Thomas Morgan's laboratory, in California. It is one of the early developments of evolutionary genetics and quantitative phylogeny, and deals with the identification and counting of chromosomal rearrangements in *Drosophila* from comparisons of genetic maps. A re-analysis by present day mathematics and computational technologies of the data produced in the 1930s unveils how the solicitation of a mathematician led the team of biologists, against their first intuition, to an erroneous conclusion about the presence of phylogenetic signal in gene orders. This illustrates the role of errors in scientific activities, as well as some unexpected difficulties of multi-disciplinary collaborations. Also underlying is the possible influence of computational complexity in understanding the directions of research in biology.

Keywords history of biology · evolutionary genetics · chromosomal inversion · genetic maps · statistics · computational complexity · scientific errors · history of interdisciplinary studies · Drosophila

Eric Tannier
Inria
Univ Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Évolutive UMR5558, F-69622 Villeurbanne, France
Tel.: +33-(0)426234474
Fax: +123-45-678910
E-mail: eric.tannier@inria.fr

This is the first time in my life I believe in constructing phylogenies, and I have to eat some of my previous words in this connection. But the thing is so interesting that both Sttt [Sturtevant] and myself are in a state of continuous excitement equal to which we did not experience for a long time.

Theodosius Dobzhansky, letter to Milislav Demerec 1936

I am rather surprised to find myself figuring out hypothetical phylogenies for the Drosophila species, and taking them more or less seriously — after all the uncomplimentary remarks I've published about such procedures.

Alfred Sturtevant, letter to Otto Mohr 1939

These two epigraphs testify to a renewed excitement for phylogeny in the first half of the twentieth century. Such signs of enthusiasm linked to the revival of this old discipline were common. They are due, among other possible explanations, to the use of cytological and genetic comparisons, offering direct access to hereditary material, on the one hand, and the use of quantified methods, often associated with an approach of objectivity, on the other. On the one hand, cytology was like "looking at the cellar window" for Anderson (1937), and is "evidence as to the germplasm itself and is, therefore, of more fundamental importance than the mere architecture erected by the germplasm itself." Chromosomes provided "high-powered morphology" for Turrill (1938). "The chromosomes are the determinants of characters", and "one cell is sufficient for the identification of the species" for McClung (1908). "Were our knowledge of cell structure in the grasshopper complete enough we might erect a system of classification based upon cytological characters, just as reasonably as we have designated one using external anatomical structures" (McClung, 1908). On the other hand, the quantified comparisons allowed by precipitin reactions (Strasser, 2010b) made Boyden (1934) write that "The fact that naturalists of recent times have so often forsaken the study of phylogeny is due more to the feeling that such a study is likely to yield little certain progress than to the belief that the problems of phylogeny are unimportant or sufficiently well analyzed.".

Of course, the conjunction of the use of "semantic"[1] characters and quantification, driven by the development of sequencing techniques and computers, has been fully realized only in the 1960s by the founders of Molecular Evolution (Suárez-Díaz, 2009; Dietrich, 2016). However the evolutionary genetics program in Thomas Morgan's laboratory beginning in 1914, and afterwards mainly led by the enemy brothers Alfred Sturtevant and Theodosius Dobzhansky, had similar epistemological characteristics[2].

The objective of this article is to give an account of a particular moment of this research, focusing on Sturtevant's attempts, during several years and with several successive PhD students, to quantify the number of inversions between two *Drosophila* homologous genetic linkage groups. Some aspects of this research, in particular a collaboration with the mathematician Morgan Ward

[1] That is, the ultimate hereditary material, directly transmitted, and not one of its products, according to the vocabulary of Zuckerkandl and Pauling (1965), see also Dietrich (1998)

[2] Despite crucial differences in the biological objects have been also described (Darden, 2005)

from Caltech and the errors that have resulted from it, have been overlooked by historical accounts of chromosome evolution studies (Hagen, 1982, 1984; Kohler, 1994; Gannett and Greisemer, 2004; Smocovitis, 2006) or of quantification tendencies in biology (Hagen, 2003; Suárez-Díaz and Anaya-Muñoz, 2008; Suárez-Díaz, 2010; Hagen, 2010). The solicitation of mathematics and mathematicians in biology was common in the 1930s, particularly in evolution. It is even part of the construction of modern synthesis (Bowler, 2003). However this collaboration has unusual aspects, because of the type of mathematics involved, which was not the one available to evolutionists, developed for example by Fisher, Wright or Haldane in statistics and population genetics. Retrospectively computational aspects are visible, handled then by underlying systematically applied algorithms. Some of the questions addressed then were solved 50 years later, or remain unsolved today. The difficulties mathematicians still encounter today to cope with the same problems are visible in Sturtevant and Ward's attempts. In particular, trying to solve the same questions myself with today's mathematics and technologies, I point to three computational and numerical approximations first made consciously by biologists, which strangely turned into mistakes and a wrong biological interpretation following the collaboration with Ward. This curious case of a an unfortunate multi-disciplinary collaboration exemplifies the presence and the importance of errors in the practice of science. We could also see an example of the often overlooked influence of computational intractability (Papadimitriou, 1993) in a biological research program.

In a first part of this manuscript, I give contextual elements concerning the use of chromosomes for evolutionary studies, worldwide and in Thomas Morgan's lab, in order to account for both the continuity and the originality of Sturtevant's research. Then in a second part I relate how Sturtevant progressed from the first genetic map to the challenge of counting inversions. in the course of the narration I give a solution to the described problems according the current state of science. I then discuss in a third part what this exercise can teach to us on the presence of errors and complexity in computational biology, as well in this early attempt as today.

1 Chromosomes as documents of evolutionary history

In the first half of the twentieth century, the development of genetics and cytology has engaged several researchers and research teams in comparative studies of chromosomes or linkage groups in order to establish evolutionary relationships and evolutionary histories (Hagen, 1982). The activity of delimiting and classifying species or making phylogenies began to include as possible markers chromosome numbers, shapes, sizes, behavior during the cell cycle, centromere positions or gene arrangements. To cite only a few landmarks, at the International Zoological Congress of 1907, in Boston, cytologist Clarence Erwin McClung stated that a character measured within the cell, such as the number of chromosomes, could be considered as informative as any morpho-

logical character for phylogenetic classification (McClung, 1908). In Berkeley from 1915 onward, plant geneticist Ernest Brown Babcock gathered a team to work on the evolution of the *Crepis* flowering plant and participated in the foundation of the "Bay Area Biosystematists" (Hagen, 1984; Smocovitis, 2009), an influential multidisciplinary group working on plant systematics. In 1926, the International Congress of Plant Science held a joint session of taxonomists, cytologists and geneticists (Hagen, 1984). In 1937, the field was sufficiently established for Edgar Anderson, from the Missouri Botanical Garden, to write an extensive review on the contribution of cytology to taxonomy in botany (Anderson, 1937). In 1938, Babcock and his collaborator George Ledyard Stebbins Jr, who would subsequently follow this research line (Smocovitis, 2006), published the influential book *The American Species of Crepis*, in which all the genetics and cytological knowledge was harnessed to decipher the complex evolutionary relationships among the *Crepis* genus (Babcock and Stebbins, 1938; Smocovitis, 2009).

In the genetics laboratory of Thomas Hunt Morgan, at Columbia from 1914 to 1928 and at Caltech thereafter, a comparable research program was engaged on *Drosophila*, the traditional model organism on which genetics was first developed (Kohler, 1994). It has been initiated by Charles Metz, born in 1889, arrived in Thomas Morgan's laboratory at Columbia in 1912, where he became interested in cytology. Metz soon remarked that his observation of *Drosophila* chromosomes in anaphase possibly carried phylogenetic information because different species had different chromosomal conformations. Combining the presence or absence of microchromosomes and the fissioned or fused state of two autosomes from 12 *Drosophila* species, Metz managed to classify chromosome organizations into five types. The types were then organized into a tree, where the branches could be interpreted as evolutionary events (Figure 1).

In his 1914 article, from which Figure 1 is reproduced, Metz speculated that the differences in the chromosome types "may indicate an evolution of chromosomes in the genus" (Metz, 1914). However, in subsequent articles on the description of chromosome types, Metz became more and more cautious about the evolutionary interpretation (Kohler, 1994), mainly because of the difficulty in assessing the homology[3] between chromosomes via this technique of independent observation in different species. This led his subsequent publications (Metz, 1916, 1918) to seem more like an organized catalog of chromosome types, with fewer evolutionary implications.

Then began the search for a technique to assess homology. One method was to produce hybrid species and observe coupled chromosomes during segregation, but had little success with *Drosophila* species (Kohler, 1994). Hybrids were common but almost always sterile. Two subsequent techniques would better allow for the possibility of assessing homology and were explored in

[3] Homology in the sense of "common evolutionary origin" was not so common in the beginning of the XXth century. The terminology was discussed and ranged from "allelomorph" to "corresponding". I most often use the current terminology for unification purposes.

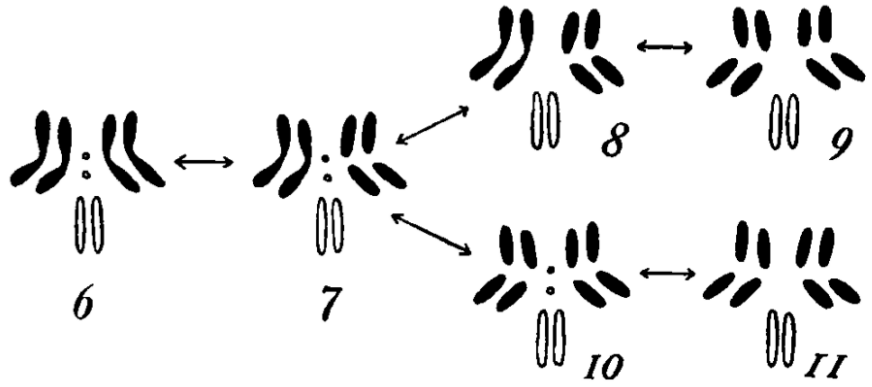

Fig. 1 Reproduced from Figure 1 in Metz (1914). Five different karyotypes, found in 12 different *Drosophila* species, are organized into a tree with a wishful evolutionary interpretation. Nodes 9 and 11 represent the same type of chromosome organization, meaning that the two phylogenetic positions are equally possible. Reproduced with the kind permission of Wiley and the *Journal of Experimental Zoology, Part A: Ecological Genetics and Physiology*.

Thomas Morgan's laboratory: mapping genes on chromosomes (from 1917) and the hybridization of polytene chromosomes (from 1936).

Charles Metz himself left Columbia for Washington in 1914 and did not participate further in activities in Columbia, even though he became an eminent *Drosophila* geneticist. However, while he was at Columbia he did not work alone and his research program continued in the hands of others. As witnessed by an acknowledgment in his 1914 article (Metz, 1914) he benefited from the help of a young student from Columbia, Alfred Sturtevant.

2 Alfred Sturtevant and Comparative Genetic Mapping, 1921 to 1941

2.1 Genetic maps and the prediction of inversions

Sturtevant, born in 1891, completed his doctorate in 1914 with Thomas Morgan at Columbia University. One of his legendary achievements was to respond to Morgan's remark that the strength of the genetic linkage between genes was possibly related to the physical distance between the genes in a chromosome. From this idea, Sturtevant defined a genetic distance as the percentage of crossing-over between two genes observed from the frequency of associated phenotypes in *Drosophila ampelophila*[4]. As this distance was close to a linear function, it was possible to position genes on a line. This led to the first ge-

[4] Renamed *melanogaster* a short time after.

netic map, placing six genes on the "sex-linked" linkage group[5] (Gannett and Greisemer, 2004).

Sturtevant's map was followed by several others issued from the same team. In particular, Morgan and Bridges' 36-marker map of the X-chromosome of *Drosophila melanogaster* (Morgan and Bridges, 1916) provoked a controversy with William Castle and, by implication, among several other researchers, about the relevance of the linear model in depicting chromosomes (Castle, 1918; Sturtevant et al., 1919; Castle, 1919; Morgan et al., 1920). Even though each protagonist gave the impression of standing firm on his respective position, the controversy clarified a great deal of the theory, as well as its underlying and *ad hoc* hypotheses.

The real starting point for evolutionary genetic studies was the discovery of a mutation of that linear structure. Inversions, which consist of replacing a chromosome segment by its reverse order, were hypothesized by Sturtevant (1921), based on the observation of differences in the arrangements of five "corresponding"[6] markers of chromosome 3 of *Drosophila simulans* and *Drosophila melanogaster*. The inversion hypothesis was confirmed by adding markers, while the comparative mapping of Sturtevant and Plunkett (1926), as depicted in Figure 2, presents a striking visual argument for it[7].

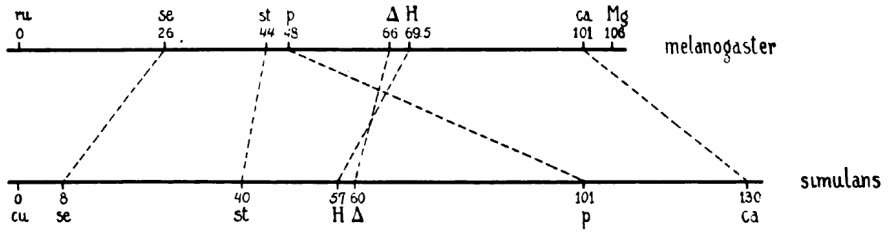

Fig. 2 Reproduced from Figure 1 of Sturtevant and Plunkett (1926): a visual argument for the existence of inversions and their utility for taxonomy. Linkage group 3 of *Drosophila simulans* (lower solid line) and *Drosophila melanogaster* (upper solid line) are compared. They consist in placing genes (points) on the line (the chromosome, or linkage group). Homology of genes is represented by dashed lines. Reproduced with the kind permission of *The Biological Bulletin*.

[5] Later named the X-chromosome in order to emphasize its oddity. The links between chromosomes and linkage groups were already well established, as witnessed by the natural use of "chromosome" in genetic studies from the 1910s.

[6] *I.e* homologous, see the previous footnote. Homology was deduced from similarity of phenotype variations during crossing experiments.

[7] Several types of translocations, *i.e.* other mutations of the linear organization of genes along chromosomes, were predicted at the same time (Bridges, 1917; Mohr, 1919; Morgan et al., 1925) and later cytologically demonstrated (Muller, 1929; Dobzhansky, 1930). They were seen more as "deficiencies", or abnormalities of karyotypes, possibly obtained under mutagenic conditions. Inversions were oppositely immediately seen as evolutionary patterns susceptible to being used in differentiating species, thus as a character for taxonomy. Translocations were later used in plant taxonomy by Babcock and Stebbins (1938).

Inversions themselves had the same status as linkage groups, that is, they were theoretical objects independent of any direct cytological observation. A cytological demonstration of their existence would be made later with the techniques of Painter (1933).

From this possibility of detecting a mutation by comparing genetic maps, Sturtevant developed a comprehensive research project in line with his collaboration with Metz. It consisted in mapping the genes of different *Drosophila* species, assessing the homologies between these genes and, from the chromosome structure, reconstructing the evolution of the associated organisms (Kohler, 1994). This research plan was not fully realized, although several publications and many unpublished results[8] confirm decisive advances and reveal some challenges.

2.2 The statement of a mathematical problem

Among the challenges was the detection of several successive overlapping inversions. Comparing two arrangements differing by one inversion was easy. However if several overlapping inversions may have occurred, which is likely if more distant species are compared, an additional difficulty arose. In 1937, Sturtevant, published with C. C. Tan, a Ph.D. student of himself and Dobzhansky, a comparison of the arrangements of 38 genes along all chromosomes of *Drosophila melanogaster* and *Drosophila pseudoobscura* (Sturtevant and Tan, 1937). The comparative maps, inferred from the homologies of genes deduced from similar phenotypic effects, are reproduced in Figure 3. Inversions are not as visible as in Figure 2 because the species are more distant, thus the accumulation of inversions has blurred the signal.

It is useful to carefully examine in Figure 3 both the organized description of the data and the paragraphs from the 1937 article by Sturtevant and Tan. Two mathematical problems are stated. One concerns the calculation of the number of successive inversions that are necessary[9] to transform a series of letters (the gene order in *melanogaster*) into alphabetical order (the gene order in *pseudoobscura*). The other asks if this number is likely to be obtained if the arrangement of letters is supposed random. For example, the sequence of chromosome IIL can be transformed into alphabetical order by two inversions as follows:

$$DEFACB \rightarrow AFEDCB \rightarrow ABCDEF.$$

The first inversion concerns the segment $DEFA$, and the second the segment $FEDCB$. The problem of computation of this number becomes tricky when genes and inversions become numerous. Notice, in the paragraph reproduced

[8] Examined by Kohler (1994), who writes that the unpublished part is of wider significance.

[9] That is, a number as small as possible, which does not invoke unnecessary inversions, *i.e.* a parsimony argument, of the same kind as the one on sequences later proposed by Camin and Sokal (1965). This number has been subsequently named the *inversion distance* of a permutation (Fertin et al., 2009).

legitimate. If the *pseudoobscura* sequence in each arm is arbitrarily taken as an alphabetical one (A B C. . .), then the *melanogaster* sequences become:

	X	L	H	F	E	B	A	D	C	K	I	J	G	M	(7)
II	*L*	D	E	F	A	C	B	(2)						
II	*R*	A	C	E	B	F	D	(4)						
III	*L*	C	F	E	B	A	D	(3)						
III	*R*	A	E	B	C	F	D	G	(3)					

The numbers in parentheses represent the numbers of successive inversions necessary to turn these sequences into alphabetical ones (in the case of *X* we are not yet certain that six inversions may not be sufficient). The mathematical properties of series of letters subjected to the operation of successive inversions do not appear to have been worked out, so that we are so far unable to present a detailed analysis. It does appear, however, that the five arms (taken together) are definitely more alike in the two species than could result from chance alone.

Fig. 3 Extracted piece from Sturtevant and Tan (1937). Letters in the far left column are chromosome names. Other capital letters are gene names. number in parentheses are minimum number of inversions, necessary to transform the arrangement of letters in a line (*melanogaster arrangement* into the alphabetical order (*pseudoobscura* arrangement). In the paragraph following the array of letters, a working program for mathematicians and (not yet existing) computer scientists is proposed (see text for details). Reproduced with permission from *Springer*.

in Figure 3, that Sturtevant and Tan recognized that their best scenario had seven inversions, but they were uncertain whether six was impossible. No detail is given about their method to find the scenario with seven and the reason why they doubt it is the minimum number but it is not hard to imagine a solution of algorithmic nature, enumerating many scenarios, that faced a too high numer of possible scenarios. The modesty of their statement was retrospectively a good intuition, since

$$LHFEBADCKIJGM$$

$$\rightarrow ABEFHLDCKIJGM$$

$$\rightarrow ABCDLHFEKIJGM$$

$$\rightarrow ABCDEFHLKIJGM$$

$$\rightarrow ABCDEFHGJIKLM$$

$$\rightarrow ABCDEFGHJIKLM$$

$$\rightarrow ABCDEFGHIJKLM$$

is one of the several possible *bona fide* sequences of six successive inversions[10]. The "detailed analysis" called for by Sturtevant and Tan (see Figure 3) would have to wait several decades before it became possible with the help of new mathematical and computational techniques (Fertin et al., 2009). Besides its anecdotal value however, this mention that the result was unsure is important, because later on it was turned into an error: meaning that the modest statement of ignorance was forgotten, while the number seven was taken for granted. Together with two other approximations concerning the statistical statement, this would lead to erroneously changing the biological conclusion.

This brings us to the second mathematical problem, the statistical one. Indeed, the last reproduced sentence in Figure 3 states that the permutations of genes, as observed in the comparative arrangement in *pseudoobscura* and *melanogaster*, "are definitely more alike in the two species than could result from chance alone." The term "definitely" is interesting for our purpose, because it illustrates the progressive extent of the quantification. It is a strongly asserted although intuitive statement that became a statistical hypothesis in a follow-up paper of 1941 by Sturtevant and another student, Edward Novitski (Sturtevant and Novitski, 1941). Novitski, like Tan before him, was first a student of Dobzhansky and continued with Sturtevant after their dispute (Novitski, 2005). He then worked on the same subjects, chromosomes and evolution, using a different approach. Together with Sturtevant, he provided a large catalog of homologies, some from the literature and some newly obtained via classical genetic techniques, along with a deeper mathematical analysis of the 1937 data. Again, a close look at the mathematical technique helps to understand the progressive introduction of quantification, and how, if it gives the impression of objectivity and fights a "methodological anxiety" (Suárez-Díaz and Anaya-Muñoz, 2008), it is not necessarily a guarantee of greater veracity or precision.

2.3 The call for a professional mathematician

Going back over the statements of Sturtevant and Tan (those reproduced in Figure 3), Sturtevant and Novitski announced that they had solicited the help of Morgan Ward, a renowned mathematician from Caltech. Sturtevant himself

[10] It is also possible to prove that six is necessary, that is, five is not possible. Here is the simplest proof I found (unfortunately involving a little bit of theory of breakpoint graphs for which one can refer to Fertin et al. (2009)): there are nine "breakpoints" in this permutation, that is, nine positions which have to be used as inversion extremities (positions between two non consecutive letters and at the extremities when the first or last letter is not properly placed according to the alphabetical order). As one inversion touches at most two breakpoints, it is necessary to have at least five distinct inversions to access nine breakpoints. If there is a scenario with five inversions, at least three of them have to be the only ones touching their breakpoints because only one re-use is allowed. This involves the existence of at least three disjoint small alternating cycles (size 4) in the breakpoint graph. The breakpoint graph of this permutation has 3 small cycles EBDC, LHKI and LHGM (found by manual enumeration of all possible small paths), which can make only two disjoint ones.

was understanding reasonably well mathematics, and Novitski (2005) retrospectively praised his "mathematical mind" compared with Dobzhansky, but he probably felt that no easy technique could solve this question and logically solicited an expert. In order to state whether the difference in arrangements was indeed "more alike [...] than could result from chance alone", calculations of means and standard deviations of inversion distances of random permutations were performed by the complete enumeration of permutations up to size six, by random sampling of 60 and 40 permutations for sizes eight and nine, and by a linear interpolation for higher numbers. Indeed, as they admitted, "For numbers of loci above nine the determination of this minimum number proved too laborious, and too uncertain, to be carried out" (Sturtevant and Novitski, 1941).

They obtained a mean of 7.6 inversions for 13 genes (see Figure 4), and concluded that, in contrast to their first intuition, "Evidently the two species are not more alike than could easily result from chance alone". The comparison of the term "definitely" in the previously quoted sentence, and the term "evidently" used here, calls for several remarks. Firstly, the second states that the first was evidently a wrong intuition, which tells us something about the scientific personality of Sturtevant: he does not hesitate to admit a supposed error from himself in strong terms. Then, if the second corrects the first by a quantitative assessment of the intuition behind the first, we can note that it largely retains an intuitive part. The authors, after having considered that the differences were "definitely" significant without having calculated them, regard it as "evidence" that seven is not significantly different from 7.6, but the argument for it depends on a standard deviation upper bound estimation (less than 1, according to the authors). A final remark is that, if we carefully check the computations, it is unfortunately noticed that the first intuitive argument, stated using this word "definitely", was correct, and that the correction, stated using the word "evidently", was not. It is sad to note that the elegance of Sturtevant, contradicting his own result, was itself a scientific error, because the first result was in fact better reflecting the data, according to their own criteria.

It is striking that, even today, no better technique is known for calculating these numbers. Only the improved performance of computers allows the modern researcher to enumerate all permutations and their inversion distance up to 13 genes in 1995 (Galvão and Dias, 2015)[11] instead of six in 1941, and an asymptotic bound for the mean has been calculated (Bafna and Pevzner, 1996) with no applicability to such small values. Hence, I have used the enumeration to compute, according to modern techniques and knowledge, the values for the numbers considered by Novitski, Sturtevant and Ward. I consider these values more precise than theirs, because I use a complete enumeration of the space instead of an extrapolation. Of course the values are the result of my

[11] It is a coincidence that the maximum number published in 1995 is precisely the one with which struggled the biologists in 1937. The similarity of the orders of magnitudes of our ability to handle these data over time is indicative of the inherent computational complexity of the problem.

Evidently the two species are not more alike than could easily result from chance alone.

TABLE 4

Comparison of the required and calculated numbers of inversions to change the melanogaster into the pseudoobscura sequences.

ELEMENT	A	B	C	D	E	TOTAL
Loci	13	6	6	6	7	
Inversions required	7	2	4	3	3	19
Inversions calculated	7.6	3.0	3.0	3.0	3.7	20.3

Fig. 4 Extracted piece from Sturtevant and Novitski (1941). The numerical errors (or imprecisions) in column A (the two computed values of 7 and 7.6 are incorrect, and should be 6 and 7.9), partially originating from the one depicted in Figure 3 led to the wrong conclusion concerning the homology between gene orders, as shown just above the table on this figure. Reproduced with the kind permission of the *Genetics Society of America* and the journal *Genetics*.

own understanding of the problem and I cannot discard the hypothesis that a future work will refute them, but I think this is the best which can be achieved with today's knowledge and technology. This modern analysis shows, for 13 genes, a mean of 7.9 instead of the interpolated value of 7.6 from 1941 (see Figure 4) and a standard deviation of 0.85 instead of the "less than one" of the 1941 estimation.

2.4 The addition of errors

The 1937 and 1941 estimations were close to the corrected values[12]. However, the small differences add up to three small errors or approximations: one in the minimum number (seven instead of the six inversions necessary in reality), one in the average number (7.6 instead of 7.9 in reality) and one in the standard deviation (less than 1 instead of 0.85). All in all these change the conclusion. After a corrected calculations, six inversions can be considered different from 7.9, with a standard deviation of 0.85[13].

[12] Note that the corrected values I give use nothing other than the published data and the statistical test proposed by the authors. Nevertheless this analysis requires computational tools that were not available at the time. There would probably be a lot more to discover if we were to redo such analysis with modern-day data.

[13] A *bona fide* statistical test here would require a p-value rather than a comparison of standard deviations. It was not imagined in the 1937 and 1941 articles but it is possible to compute empirical p-values from samples from 1,000 random uniform permutations. This gives an approximate probability of 0.06 of achieving six inversions or fewer for 13 genes, a probability 0.2 of achieving two inversions or less for six genes and a probability 0.35 of achieving three inversions or less for seven genes. Each chromosome taken independently is hardly conclusive. Together, they can be considered significantly far from random according to this measure and the usual significance thresholds.

It is retrospectively vertiginous that Sturtevant and Novitski (1941), assisted by a mathematician, claimed in 1941 to correct the intuitive statement of Sturtevant and Tan (1937) in 1937[14], but did not correct the real error that was present in the 1937 paper (the number of inversions necessary to sort the letters in alphabetical order). Instead they corrected, in the wrong direction, the statement that the permutations were different from random ones, which introduced two errors while failing to correct the only real one. The help of a mathematician, which ordinarily would have been considered a good idea for such a problem, was disastrous in this case, as it undermined, for the wrong reasons, the sound intuitions that the biologists first had.

Detecting these errors goes beyond the mathematical exercise and an exaggerated attention to the details. And it is not only the story of a an unusual and unfortunate early collaboration of biologists with a mathematician. Success stories are more often related than errors, but it could be that the fate of scientific research is sometimes influenced by mistakes of diverse types (Firestein, 2015; Livio, 2014). Here, the fact that very closely related species, such as *melanogaster* and *simulans*, had apparently not conserved any detectable similarity in gene orders might have participated to orient genetic research in other directions. Indeed, a *Drosophila* phylogeny based on chromosomes was hardly conceivable according to this conclusion.

Little changed in that direction after 1941. In a 50-page landmark article about a *Drosophila* genus phylogeny, Sturtevant (1942) included only two pages on chromosomes, from which he derived no decisive phylogenetic relationship. The document is mainly based on morphological characters. This can be contrasted with the book of Babcock and Stebbins (1938), which acknowledged the possible use of chromosomes for the phylogeny and evolutionary history of *Crepis* , even though it was biologically more complex because of hybridization and the diversity in the modes of reproduction of plants. Babcock started his research on *Crepis* with a wish to find an equivalent of *Drosophila* in plants, to explore to what extent the results of Morgan's fly laboratory were generalizable (Smocovitis, 2009). He did not fully succeed in this precise goal but in some aspects went beyond the *Drosophila* research in evolutionary genetics and cytology.

2.5 Epilogue: Theodosius Dobzhansky and Polytene Chromosomes

Of course the evolutionary genetics project Metz and Sturtevant had in 1914 has not stopped because of a few mathematical mistakes made in the 1930s. One important reason is that it necessitated a prohibitive amount of work to assess the homology between genes and between chromosome segments.

[14] They insisted and elaborated on the "incorrect conclusion of Sturtevant and Tan", which was actually reasonable from what I can judge today given the data. They also insisted that, in hindsight, the only retrospectively wrong statement in the 1937 article was in fact correct: "it may be noted that this revision does not change the number of inversions required to transform one sequence into the other", thereby transforming the modest imprecision into an error.

In the studies reported here, from 1937 and 1941, a catalog of homologous genes was gathered according to the similarities in phenotype variations. This tedious method, difficult to automate, could not be envisaged beyond a certain evolutionary depth. A new hope in that direction was introduced with the discovery of a technique for comparing polytene chromosomes.

In 1933, Theophilus Painter invented an alternative technique (Painter, 1933, 1934) to detect similarities and differences between chromosome organizations. It consisted in hybridizing polytene chromosomes from different strains or species of insects and visualizing chromosome inversions (Figure 5). Immediately, geneticists and cytologists exploited the potential of this technique from a comparative and evolutionary perspective (Gannett and Greisemer, 2004). It has remained a widely used technique for studying rearrangements until today.

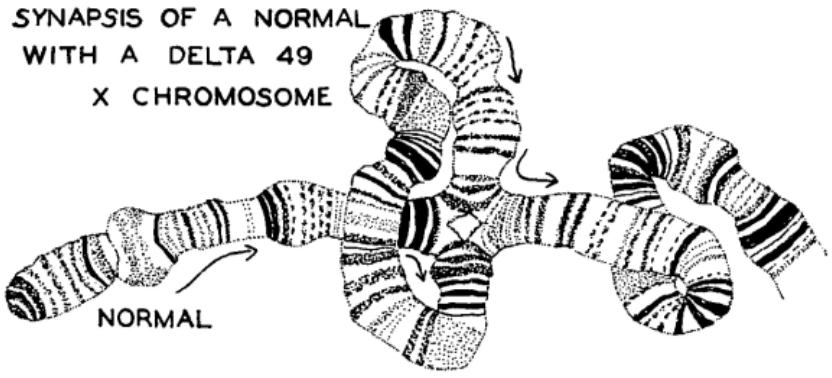

Fig. 5 Concatenated polytene X-chromosomes of *Drosophila melanogaster*, from Figure 1 in Painter (1934). An inversion on the "delta 49" variant strain is made visible by a loop on the concatenated giant chromosome. Reproduced with the kind permission of the *American Genetics Association* and the *Journal of Heredity*.

In particular, it was used by Theodosius Dobzhansky, first in association with Sturtevant and then independently after their partnership ended. Dobzhansky was born in 1901 and arrived at Morgan's laboratory at Columbia in 1927. He then moved to Caltech with the group in 1928. At first, he and Sturtevant had a very close relationship as collaborators and friends. Dobzhansky used to say that he owed his life to Sturtevant, who had strongly supported him in securing a position in the United States.

Dobzhansky and Sturtevant differed in terms of practice and vision. While their differences initially complemented each other, they soon caused a problem in their relationship. Dobzhansky saw himself as a field naturalist interested in natural populations, while Sturtevant worked with species grown in the laboratory (Kohler, 1994). Dobzhansky was interested in the theoretical processes of evolution, while Sturtevant was more of a geneticist and systematicist.

Dobzhansky published copiously and was expansive, extravagant, irreverent, religious and sensitive to celebrity: in short, the opposite of Sturtevant's personality (Novitski, 2005).

Initially their differences yielded very productive and innovative scientific research. Dobzhansky collected many *pseudoobscura* strains from all around California and the United States[15] and promoted the cytological studies, while Sturtevant took what was interesting for genetics and, in particular, his now old project of comparing chromosomes from an evolutionary perspective. The field strains collected by Dobzhansky and the Painter technique seemed ideal for studying the evolutionary history from chromosome inversions. Indeed, the two identified races of *Drosophila pseudoobscura* (see Figure 6) showed some intraracial variability. It was possible to classify all the strains according to the structure of their chromosomes, such that a hypothetical evolutionary history could be deduced from this classification. In 1936, the first phylogenetic tree whose branches are chromosome inversions was published (Sturtevant and Dobzhansky, 1936) (see Figure 6).

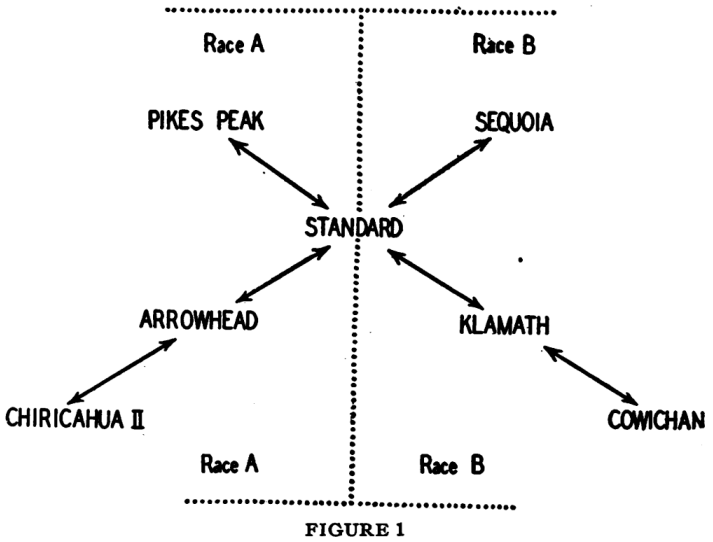

Fig. 6 A phylogeny of seven *Drosophila pseudoobscura* strains, from Sturtevant and Dobzhansky (1936). Reproduced with the tacit permission of the *National Academy of Science* of the United States of America.

An augmented phylogeny was published in 1938 by Dobzhansky and Sturtevant, including 17 *pseudoobscura* strains (Dobzhansky and Sturtevant, 1938), although the context was already different. The complementarity of the two

[15] When they both discovered a new species *Drosophila miranda*, Dobzhansky first named it *sturtevantiana*.

authors had turned into opposition. The retirement of Morgan, aged 70 in 1936, and the inevitable succession, revealed ambitions, frustrations and fears (Kohler, 1994). It prompted the beginning of the rift. The 1938 paper, the last of their fruitful collaboration, was mainly Dobzhansky's work according to himself (in his oral history interviews of 1962, cited by Kohler (1994)), a situation which continued in the series "Genetics of Natural Populations" (Dobzhansky and Queal, 1938). In 1940, Dobzhansky left Caltech for Columbia[16].

Compared to genetic maps, assessing homology using the cytology of polytene chromosomes was fast and much less costly, which partly explains its immediate and long-lasting success. Regarding longer-term evolution, however, it was somewhat also limited. While it was extremely efficient at comparing different natural strains of a single species, it could not be used to compare two distant species in which chromosomes are separated by several inversions. If there were more than three overlapping inversions on the same chromosome, the technique yielded almost no interpretable observations. Intermediary steps were required.

However Dobzhansky and Powell (1975) followed by others (Carson and Kaneshiro, 1976) finally established the phylogeny of *Drosophila* clades with more than a hundred arrangements and several hundred inversions. Polytene chromosomes are still used in insect chromosome comparisons, and cytology has turned into cytogenetics, with extremely productive follow-ups (Carson and Kaneshiro, 1976; Brehm, 1990; Dutrillaux and Dutrillaux, 2012).

It is noticeable that, when the comparison of protein sequences was made possible by new molecular techniques (Zuckerkandl and Pauling, 1965), Dobzhansky did not engage in this field. Instead, he attacked this new trend in molecular studies by taking a "naturalist's" point of view (Dobzhansky, 1963, 1966; Dietrich, 1998). This is symptomatic of how genetics in the first half of the twentieth century on one side, and molecular evolution in the second half on the other, have negotiated their place in evolutionary biology (Hagen, 1999; Dietrich, 1998; Sommer, 2008). Dobzhansky, who could accommodate genetics as a component of the study of evolution and participate in the unification of genetics and evolution, could not accept the imperialistic statements of Zuckerkandl on protein comparisons, claiming that proteins were possibly the only reliable materials with which to study evolution.

3 Discussion

3.1 Counting mutations as a computational biology problem

The introduction of measures, statistics and mathematics into evolutionary studies and phylogeny has traversed all the twentieth century (Hagen, 2003; Sommer, 2008; Suárez-Díaz and Anaya-Muñoz, 2008; Suárez-Díaz, 2010). This

[16] The story of the rift between Dobzhansky and Sturtevant is extensively related in several places, for example, Kohler (1994); Novitski (2005).

tendency is visible in biology, science and society in general (Kay, 1993; de Chadarevian and Kamminga, 1998; Porter, 1996). While it was not systematically the case, several researchers valued this use as a possibility to turn phylogeny into a *bona fide* science.

Computing evolutionary distances has been an important activity for establishing phylogenetic relationships. In the first half of the twentieth century it was done for example with serological and immunological reactions (de Chadarevian, 1996; Strasser, 2010b; Hagen, 2010), DNA hybridization (Suárez-Díaz, 2014), which results hopefully would reflect the amount of divergence between proteins or chromosomes.

It has been observed that the kind of quantification involved has become a direct quantifying of DNA mutations only with the birth of molecular biology in the 1960s (Hagen, 1999, 2003, 2010; Strasser, 2010b; Suárez-Díaz, 2014; Dietrich, 1994, 1998; Morgan, 1998; Sommer, 2008), involving a particular type of mathematics, often aided by the use of computers (Hagen, 2000, 2001; Strasser, 2010a).

The kind of mathematics used by Sturtevant and Ward for counting inversions has a very special status to this respect. On the one hand, it consists in counting successive mutations in semantic characters, just like counting substitutions in protein or gene sequences. In that sense, it is more related to the mathematics developed in the 1960s than the quantification from the 1930s developed as a proxy, "waiting for sequences" (Hagen, 2010). On the other hand, counting inversions has a decisive difference with counting point mutations: as an approximation, point mutations can be considered independent from each other, while overlapping inversions are inaccessible to this kind of simplifying hypothesis. That can explain why, even if introduced 30 years before, counting inversions is until today much less developed than counting point mutations.

The technique for counting inversion involves algorithms. In the absence of a trivial mathematical formula for estimating a number of inversions, Sturtevant, his students and Ward must have applied some method. They have counted inversions between several hundreds of pairs of permutations: not only those coming from their biological data, but also the whole set of permutations up to 6 elements, plus samples of permutations of size seven to nine. They do not detail how they achieved this but admit "For numbers of loci above nine the determination of this minimum number proved too laborious, and too uncertain, to be carried out". This means that they were certain for numbers of loci up to nine, which is already a difficult exercise on some permutations. We do not know how they arrived to this confidence but we can only imagine an automatic method, that is, an algorithm.

Algorithms were not at all common to solve biological problems. They were known to mathematicians but their usage was not formalized. The famous Turing articles are contemporaneous from Sturtevant's researches (Turing, 1936). It is by the way singular that this mention of what we can call today a computational biology problem originates from the laboratory of Thomas Morgan, to whom legend attributes an aversion to computers. It is said that he banned

Friden calculators from the biology department at Caltech, because he mistrusted all quantitative and automatic results[17].

Even if the possibility of analysing DNA sequences at the whole chromosome level has brought more data, more precision and more evolutionary depth, the principle of chromosome comparison, oppositely to the detection of point mutations in genes, has not changed with the availability of sequences and still consists in counting inversions (Pevzner and Tesler, 2003; Murphy et al., 2005). The semantic stage had been reached in the 1930s for this kind of mutation.

3.2 Interdisciplinary studies

Interdisciplinarity is also an element of this history, with ambiguous outcomes. It was already not rare that mathematicians got involved in biological problems, or that biologists use statistics, or that biologists and mathematicians would collaborate (Hagen, 2003). Sturtevant was able to use basic mathematics, had a "mathematical mind" according to his student Novitski (Novitski, 2005), and began as an autonomous and modest mathematician the comparison of chromosomes (Sturtevant and Tan, 1937). Along the lines of this article, a mathematical and a computational problem are modeled, and some solutions arise. It is acknowledged that some results are "not yet certain". This might have decided Sturtevant to contact Morgan Ward, from the mathematics department of his university. Sturtevant was able to model and solve to a certain extent the mathematical problem, but was not "yet" certain of the result. He already had in mind to check it and may be he had in mind to have it checked by a professional. This interdisciplinary practice is not the kind practiced by Dobzhansky and Wright at the same time, where both equally involved, complementary. Ward is not involved in the general project, as witnessed by an acknowledgement but no co-authorship. From this collaboration with a mathematician, Sturtevant has made some advances on his problem, and also adopts a less modest style concerning the results. He trusts in Ward more than in himself concerning mathematical results. Part of this rigorous and modest mind has led him to engage in an erroneous direction.

3.3 The importance computational complexity

The errors I have reported were obviously not the result of incompetency or poor intuition on the part of researchers. They could be due insufficient interest in the problem from the practitioners. Indeed, assessing the homologies between genes by genetic techniques was time-consuming, costly, not susceptible to automation nor generalization to more distant species, which led little chances of success to a wide research program based on this technique. This might explain that the results of the comparison between *melanogaster* and

[17] This story is attributed to Charlie Munger in Belevin (2007).

pseudo-obscura have not been reproduced for other species, nor the mathematical techniques have been refined, and the errors corrected by further work.

However the facts that Sturtevant requested the help of a professional mathematician despite being himself a decent amateur, that two publications, with two different Ph-D students, with four years distance, mention the mathematical problem, that between the two, intuitive statements are abandoned for quantified statements, that a supposed error in the first publication is corrected in the second, these facts witness a reasonable interest by Sturtevant for obtaining a right answer to the problems he raised.

Thus one of the reasons why he did not achieve it at that time might be because the mathematics raised by successive overlapping inversions is intractable[18]. Not only are these problems hard[19], but the possibility of their resolution required the development of theoretical computer science and algorithmics, whose starting point can be seen as concomitant with Sturtevant's research. Although little known and often despised by most biologists, computational complexity is a constraint that can have an influence on the directions taken by biological researches, and the history narrated here could be an example of the influence of such a constraint. Computational complexity might also explain why until today some biological processes are intensively studied (point mutations in DNA) while others are much less quantified (repetitions, inversions).

3.4 The importance of errors

Good histories of scientific errors are scarce (Firestein, 2015; Livio, 2014). Their importance, their influence is often hidden by practitioners who like to put forward successes, or use errors as pedagogical scarecrows (Bosch, 2018) while historians are wary of their anachronistic character: pointing at an error is a way to include the future refutation as an element of understanding the past.

Errors are nonetheless a good way to approach the ambiguous relationship that scientists have to the truth, and can be characteristic of a knowledge state, of a sociological or psychological context. Here, the errors point at some challenges that mathematical and computational biology have to face regarding computational complexity, interdisciplinary studies, and how a number, associated with a caution note, can be taken four years later for a trustworthy value, the authors themselves forgetting the note.

[18] Note that this contrasts with the later history of protein sequence alignment, where a comparison between two related sequences was possible without excessive mathematical involvement. I do not mean that there is no interesting mathematical problem associated with sequence alignment but that it was easier to solve with the intuitive ideas of biologists than to compute an inversion distance. (see, for example, Margoliash (1963)).

[19] Finding the minimum number of inversions to transform a sequence of letters into alphabetical order is provably intractable (Caprara and Lancia, 2000).

Acknowledgements

Thanks to Istvan Miklos for having pointed to me the 1937 article by Sturtevant, and to Vincent Daubin and Bastien Boussau for having given me the occasion to construct and present part of this work at the Jacques Monod conference in 2016, "Molecules as documents of evolutionary history : 50 years after". Thanks also to several anonymous historians who have kindly helped me to improve the historical issues, even if the argument still looked too teleological and anachronistic for them after these improvements.

References

Anderson, E. 1937, Jul. Cytology in its relation to taxonomy. *The Botanical Review 3*(7), 335–350.

Babcock, E. B. and G. L. Stebbins 1938. *The American species of Crepis: Their interrelationships and distribution as affected by polyploidy and apomixis.* Washington, D.C: Carnegie Institution of Washington.

Bafna, V. and P. A. Pevzner 1996, February. Genome rearrangements and sorting by reversals. *SIAM J. Comput. 25*(2), 272–289.

Belevin, P. 2007. *Seeking Wisdom: From Darwin to Munger.* PCA Publications.

Bosch, G. 2018, February. Train PhD students to be thinkers not just specialists. *Nature 554*(7692), 277–277.

Bowler, P. J. 2003, 07. *Evolution.* University of California Press.

Boyden, A. 1934. Precipitins and phylogeny in animals. *The American Naturalist 68*(719), 516–536.

Brehm, A. 1990. *Phylogénie de neuf espèces de Drosophila du groupe obscura d'après les homologies de segments des chromosomes polytènes.* Ph. D. thesis, Université de Lyon 1.

Bridges, C. B. 1917. Deficiency. *Genetics 2*, 445–465.

Camin, J. H. and R. R. Sokal 1965, sep. A method for deducing branching sequences in phylogeny. *Evolution 19*(3), 311–326.

Caprara, A. and G. Lancia 2000. Experimental and statistical analysis of sorting by reversals. In D. Sankoff and J. H. Nadeau (Eds.), *Comparative Genomics*, pp. 171–183. Springer.

Carson, H. L. and K. Y. Kaneshiro 1976. Drosophila of hawaii: systematics and ecological genetics. *Annual Review of Ecology and Systematics 7*(1), 311–345.

Castle, W. E. 1918, Feb. Is the arrangement of the genes in the chromosome linear? *Proc Natl Acad Sci U S A 5*(2), 25–32.

—— 1919, Nov. Are genes linear or non-linear in arrangement? *Proc Natl Acad Sci U S A 5*(11), 500–506.

Darden, L. 2005, jun. Relations among fields: Mendelian, cytological and molecular mechanisms. *Studies in History and Philosophy of Science Part*

C: Studies in History and Philosophy of Biological and Biomedical Sciences 36(2), 349–371.

de Chadarevian, S. 1996, Sep. Sequences, conformation, information: Biochemists and molecular biologists in the 1950s. *Journal of the History of Biology 29*(3), 361–386.

de Chadarevian, S. and H. Kamminga (Eds.) 1998. *Molecularizing Biology and Medicine New Practices and Alliances, 1920s to 1970s*. Taylor and Francis.

Dietrich, M. R. 1994, Mar. The origins of the neutral theory of molecular evolution. *Journal of the History of Biology 27*(1), 21–59.

—— 1998. Paradox and persuasion: negotiating the place of molecular evolution within evolutionary biology. *J Hist Biol 31*(1), 85–111.

Dietrich, M. R. 2016. History of molecular evolution. In *Encyclopedia of Evolutionary Biology*. Elsevier.

Dobzhansky, T. 1930. Translocations involving the third and the fourth chromosomes of drosophila melanogaster. *Genetics 15*(4), 347–399.

—— 1963. Evolutionary and population genetics. *Science 142*(3596), 1131–1135.

—— 1966. Are naturalists old-fashioned? *The American Naturalist 100*(915), 541–550.

Dobzhansky, T. and J. R. Powell 1975. Drosophila pseudoobscura and its american relatives, drosophila persimilis and drosophila miranda. In R. King (Ed.), *Invertebrates of Genetic Interest*, pp. 537–587. Plenum Press.

Dobzhansky, T. and M. Queal 1938. Genetics of natural populations. i. chromosome variation in populations of drosophila pseudoobscura inhabiting isolated mountain ranges. *Genetics 23*(2), 239.

Dobzhansky, T. and A. H. Sturtevant 1938, Jan. Inversions in the chromosomes of drosophila pseudoobscura. *Genetics 23*(1), 28–64.

Dutrillaux, A.-M. and B. Dutrillaux 2012. Chromosome analysis of 82 species of scarabaeoidea (coleoptera), with special focus on nor localization. *Cytogenetic and genome research 136*, 208–219.

Fertin, G., A. Labarre, I. Rusu, E. Tannier, and S. Vialette 2009. *Combinatorics of Genome Rearrangements*. London: MIT press.

Firestein, S. 2015. *Failure*. oxford university press.

Galvão, G. R. and Z. Dias 2015, January. An audit tool for genome rearrangement algorithms. *J. Exp. Algorithmics 19*, 1.7:1.1–1.7:1.34.

Gannett, L. and J. R. Greisemer 2004. Classical genetics and the geography of genes. In Rheinberger and Gaudilliere (Eds.), *Classical Genetic Research and Its Legacy*, pp. 57–88. London and New York: Routledge.

Hagen, J. B. 1982. *Experimental Taxonomy, 1930-1950: The Impact of Cytology, Ecology, and Genetics on Ideas of Biological Classification*. Ph. D. thesis, Oregon State University.

Hagen, J. B. 1984, Jun. Experimentalists and naturalists in twentieth-century botany: Experimental taxonomy, 1920–1950. *Journal of the History of Biology 17*(2), 249–270.

—— 1999. Naturalists, molecular biologists, and the challenges of molecular evolution. *Journal of the History of Biology 32*(2), 321–341.

638 —— 2000. The origins of bioinformatics. *Nature Reviews Genetics 1*(3), 231.
639 —— 2001. The introduction of computers into systematic research in the
640 united states during the 1960s. *Stud Hist Phil Biol and Biomed Sci. 32*,
641 291–314.
642 —— 2003. The statistical frame of mind in systematic biology from quantita-
643 tive zoology to biometry. *Journal of the History of Biology 36*(2), 353–384.
644 —— 2010. Waiting for sequences: Morris goodman, immunodiffusion experi-
645 ments, and the origins of molecular anthropology. *Journal of the History of*
646 *Biology 43*(4), 697–725.
647 Kay, L. E. 1993. *The Molecular Vision of Life: Caltech, The Rockefeller Foun-*
648 *dation, and the Rise of the New Biology*. Oxford University press.
649 Kohler, R. E. 1994. *Lords of the fly: Drosophila genetics and the experimental*
650 *life*. University of Chicago Press.
651 Livio, M. 2014. *Brilliant Blunders: From Darwin to Einstein - Colossal Mis-*
652 *takes by Great Scientists That Changed Our Understanding of Life and the*
653 *Universe*. Brilliance Audio.
654 Margoliash, E. 1963. Primary structure and evolution of cytochrome c. *Pro-*
655 *ceedings of the National Academy of Sciences 50*(4), 672–679.
656 McClung, C. E. 1908. Cytology and taxonomy. *Kansas University Science*
657 *Bulletin 4*(7), 199–215.
658 Metz, C. W. 1914. Chromosome studies in the diptera. i. a preliminary survey
659 of five different types of chromosome groups in the genus drosophila. *Journal*
660 *of Experimental Zoology Part A: Ecological Genetics and Physiology 17*(1),
661 45–59.
662 —— 1916. Chromosome studies on the Diptera. III. additional types of chro-
663 mosome groups in the Drosophilidae. *The American Naturalist 50*(598),
664 587–599.
665 —— 1918. Chromosome studies on the Diptera. *Zeitschrift für induktive*
666 *Abstammungs-und Vererbungslehre 19*(3), 211–213.
667 Mohr, O. L. 1919. Character changes caused by mutation of an entire region
668 of a chromosome in drosophila. *Genetics 4*, 275–282.
669 Morgan, G. J. 1998. Emile Zuckerkandl, Linus Pauling, and the molecular
670 evolutionary clock, 1959-1965. *J Hist Biol 31*(2), 155–178.
671 Morgan, T. H. and C. B. Bridges 1916. *Sex-linked inheritance in Drosophila*.
672 Carnegie Inst. Washington, Publ.
673 Morgan, T. H., C. B. Bridges, and A. H. Sturtevant 1925. *The Genetics of*
674 *Drosophila*. Bibliographia Genetica.
675 Morgan, T. H., A. H. Sturtevant, and C. B. Bridges 1920. The evidence for
676 the linear order of the genes. *Proc Natl Acad Sci U S A 6*(4), 162–164.
677 Muller, H. J. 1929. The first cytological demonstration of a translocation in
678 drosophila. *The American Naturalist 63*(689), 481–486.
679 Murphy, W. J., D. M. Larkin, A. Everts-van der Wind, G. Bourque, G. Tesler,
680 L. Auvil, J. E. Beever, B. P. Chowdhary, F. Galibert, L. Gatzke, C. Hitte,
681 S. N. Meyers, D. Milan, E. A. Ostrander, G. Pape, H. G. Parker, T. Raud-
682 sepp, M. B. Rogatcheva, L. B. Schook, L. C. Skow, M. Welge, J. E. Womack,
683 S. J. O'brien, P. A. Pevzner, and H. A. Lewin 2005, Jul. Dynamics of mam-

684 malian chromosome evolution inferred from multispecies comparative maps.
685 *Science 309*(5734), 613–617.
686 Novitski, E. 2005. *Sturtevant and Dobzhansky: Two Scientists at Odds, With*
687 *a Student's Recollections*. Bloomington: Xlibris Corporation.
688 Painter, T. S. 1933. A new method for the study of chromosome rearrange-
689 ments and the plotting of chromosome maps. *Science 78*, 585–586.
690 —— 1934. Salivary chromosomes and the attack on the gene. *Journal of*
691 *Heredity 25*(12), 465–476.
692 Papadimitriou, C. H. 1993. *Computational Complexity*. Pearson.
693 Pevzner, P. and G. Tesler 2003, Jan. Genome rearrangements in mammalian
694 evolution: lessons from human and mouse genomes. *Genome Res 13*(1),
695 37–45.
696 Porter, T. M. 1996. *Trust in numbers*. Princeton University Press.
697 Smocovitis, V. B. 2006. Keeping up with dobzhansky: G. ledyard stebbins,
698 jr., plant evolution, and the evolutionary synthesis. *Hist. Phil. Life Sci., 28*,
699 9–48.
700 —— 2009, aug. The "Plant Drosophila": E. b. babcock, the GenusCrepis,
701 and the evolution of a genetics research program at berkeley, 1915–1947.
702 *Historical Studies in the Natural Sciences 39*(3), 300–355.
703 Sommer, M. 2008, Sep. History in the gene: Negotiations between molecular
704 and organismal anthropology. *Journal of the History of Biology 41*(3), 473–
705 528.
706 Strasser, B. J. 2010a. Collecting, comparing, and computing sequences: the
707 making of Margaret O. Dayhoff's atlas of protein sequence and structure,
708 1954–1965. *Journal of the History of Biology 43*(4), 623–660.
709 —— 2010b. Laboratories, museums, and the comparative perspective: Alan
710 A. Boyden's quest for objectivity in serological taxonomy, 1924-1962. *Hist*
711 *Stud Nat Sci 40*(2), 149–182.
712 Sturtevant, A. H. 1921, Aug. A case of rearrangement of genes in drosophila.
713 *Proc Natl Acad Sci U S A 7*(8), 235–237.
714 —— 1942. The classification of the genus drosophila, with descriptions of nine
715 new species. *Austin: The University of Texas Publication 4213*, 5–51.
716 Sturtevant, A. H., C. B. Bridges, and T. H. Morgan 1919, May. The spatial
717 relations of genes. *Proc Natl Acad Sci U S A 5*(5), 168–173.
718 Sturtevant, A. H. and T. Dobzhansky 1936, Jul. Inversions in the third chro-
719 mosome of wild races of drosophila pseudoobscura, and their use in the study
720 of the history of the species. *Proc Natl Acad Sci U S A 22*(7), 448–450.
721 Sturtevant, A. H. and E. Novitski 1941. The homologies of chromosome ele-
722 ments in the genus drosophila. *Genetics 26*, 517–541.
723 Sturtevant, A. H. and C. R. Plunkett 1926. Sequence of corresponding third-
724 chromosome genes in drosophila melanogaster and d. simulans. *Biol Bull 50*,
725 56–60.
726 Sturtevant, A. H. and C. C. Tan 1937. The comparative genetics of Drosophila
727 Pseudoobscura and D. Melanogaster. *Journal of Genetics 34*, 415–432.
728 Suárez-Díaz, E. 2009, Mar. Molecular evolution: concepts and the origin of
729 disciplines. *Stud Hist Philos Biol Biomed Sci 40*(1), 43–53.

730 —— 2010. Making room for new faces: evolution, genomics and the growth of
731 bioinformatics. *History and Philosophy of the Life Sciences 32*(1), 65–89.
732 —— 2014, Aug. The long and winding road of molecular data in phylogenetic
733 analysis. *Journal of the History of Biology 47*(3), 443–478.
734 Suárez-Díaz, E. and V. H. Anaya-Muñoz 2008, Dec. History, objectivity, and
735 the construction of molecular phylogenies. *Stud Hist Philos Biol Biomed*
736 *Sci 39*(4), 451–468.
737 Turing, A. M. 1936. On computable numbers, with an application to
738 the entscheidungsproblem. *Proceedings of the London mathematical soci-*
739 *ety 2*(1), 230–265.
740 Turrill, W. B. 1938. The expansion of taxonomy with special reference to
741 spermatophyta. *BioL Rev. 13*, 342–373.
742 Zuckerkandl, E. and L. Pauling 1965, Mar. Molecules as documents of evolu-
743 tionary history. *J Theor Biol 8*(2), 357–366.