



HAL
open science

The curious case of a hapless mathematical contribution to biology

Eric Tannier

► **To cite this version:**

Eric Tannier. The curious case of a hapless mathematical contribution to biology. 2021. hal-03153696v1

HAL Id: hal-03153696

<https://hal.science/hal-03153696v1>

Preprint submitted on 26 Feb 2021 (v1), last revised 20 Sep 2022 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **The curious case of a hapless mathematical**
2 **contribution to biology**

3 **Chromosome inversions in *Drosophila*, 1937-1941**

4 **Eric Tannier**

5 **Abstract** This is the story, revisited by current eyes and means, of a mathe-
6 matical biology problem explored in the 1930s in Thomas Morgan's laboratory,
7 in California. It is one of the early developments of evolutionary genetics and
8 quantitative phylogeny, and deals with the identification and counting of chro-
9 mosomal rearrangements in *Drosophila* from comparisons of genetic maps. A
10 re-analysis by present day mathematics and computational technologies of the
11 data produced in the 1930s unveils how the solicitation of a mathematician
12 led the team of biologists, against their first intuition, to an erroneous conclu-
13 sion about the presence of phylogenetic signal in gene orders. This illustrates
14 the role of errors in scientific activities, as well as some unexpected difficulties
15 of multi-disciplinary collaborations. Also underlying is the possible influence
16 of computational complexity in understanding the directions of research in
17 biology.
18

19 **Keywords** history of biology · evolutionary genetics · chromosomal inversion ·
20 genetic maps · statistics · computational complexity · scientific errors · history
21 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

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

26 Theodosius Dobzhansky, letter to Milislav Demerec 1936

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

30 Alfred Sturtevant, letter to Otto Mohr 1939

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

52 Of course, the conjunction of the use of "semantic"¹ characters and quan-
53 tification, driven by the development of sequencing techniques and computers,
54 has been fully realized only in the 1960s by the founders of Molecular Evo-
55 lution (Suárez-Díaz, 2009; Dietrich, 2016). However the evolutionary genetics
56 program in Thomas Morgan's laboratory beginning in 1914, and afterwards
57 mainly led by the enemy brothers Alfred Sturtevant and Theodosius Dobzhan-
58 sky, had similar epistemological characteristics².

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

¹ That is, the ultimate hereditary material, directly transmitted, and not one of its prod-
ucts, 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)

64 from Caltech and the errors that have resulted from it, have been overlooked
65 by historical accounts of chromosome evolution studies (Hagen, 1982, 1984;
66 Kohler, 1994; Gannett and Greisemer, 2004; Smocovitis, 2006) or of quan-
67 tification tendencies in biology (Hagen, 2003; Suárez-Díaz and Anaya-Muñoz,
68 2008; Suárez-Díaz, 2010; Hagen, 2010). The solicitation of mathematics and
69 mathematicians in biology was common in the 1930s, particularly in evolu-
70 tion. It is even part of the construction of modern synthesis (Bowler, 2003).
71 However this collaboration has unusual aspects, because of the type of math-
72 ematics involved, which was not the one available to evolutionists, developed
73 for example by Fisher, Wright or Haldane in statistics and population genetics.
74 Retrospectively computational aspects are visible, handled then by underly-
75 ing systematically applied algorithms. Some of the questions addressed then
76 were solved 50 years later, or remain unsolved today. The difficulties mathe-
77 maticians still encounter today to cope with the same problems are visible in
78 Sturtevant and Ward's attempts. In particular, trying to solve the same ques-
79 tions myself with today's mathematics and technologies, I point to three com-
80 putational and numerical approximations first made consciously by biologists,
81 which strangely turned into mistakes and a wrong biological interpretation
82 following the collaboration with Ward. This curious case of a an unfortunate
83 multi-disciplinary collaboration exemplifies the presence and the importance
84 of errors in the practice of science. We could also see an example of the often
85 overlooked influence of computational intractability (Papadimitriou, 1993) in
86 a biological research program.

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

96 **1 Chromosomes as documents of evolutionary history**

97 In the first half of the twentieth century, the development of genetics and
98 cytology has engaged several researchers and research teams in comparative
99 studies of chromosomes or linkage groups in order to establish evolutionary
100 relationships and evolutionary histories (Hagen, 1982). The activity of delimit-
101 ing and classifying species or making phylogenies began to include as possible
102 markers chromosome numbers, shapes, sizes, behavior during the cell cycle,
103 centromere positions or gene arrangements. To cite only a few landmarks, at
104 the International Zoological Congress of 1907, in Boston, cytologist Clarence
105 Erwin McClung stated that a character measured within the cell, such as the
106 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 homology³ 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.

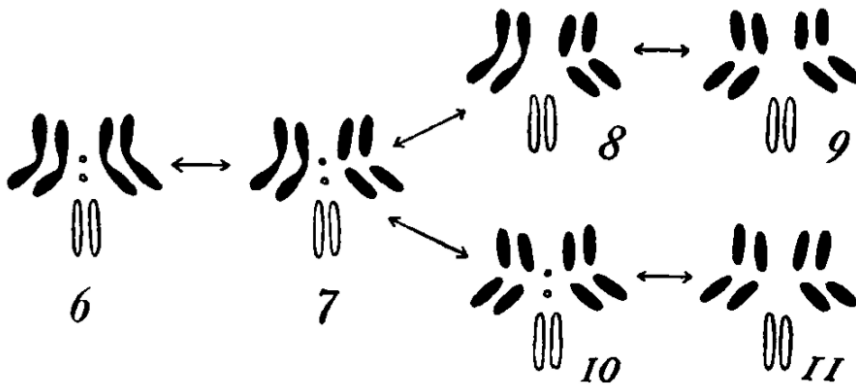


Fig. 1 Reproduced from Figure 1 in Metz (1914). Five different karyotypes, found in 12 different *Drosophila* species, are organized into a tree with a wishful evolutionary interpretation. Nodes 9 and 11 represent the same type of chromosome organization, meaning that the two phylogenetic positions are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

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

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

157 **2 Alfred Sturtevant and Comparative Genetic Mapping, 1921 to** 158 **1941**

159 2.1 Genetic maps and the prediction of inversions

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

⁴ Renamed *melanogaster* a short time after.

168 netic map, placing six genes on the "sex-linked" linkage group⁵ (Gannett and
169 Greisemer, 2004).

170 Sturtevant's map was followed by several others issued from the same team.
171 In particular, Morgan and Bridges' 36-marker map of the X-chromosome of
172 *Drosophila melanogaster* (Morgan and Bridges, 1916) provoked a controversy
173 with William Castle and, by implication, among several other researchers,
174 about the relevance of the linear model in depicting chromosomes (Castle,
175 1918; Sturtevant et al., 1919; Castle, 1919; Morgan et al., 1920). Even though
176 each protagonist gave the impression of standing firm on his respective position,
177 the controversy clarified a great deal of the theory, as well as its underlying
178 and *ad hoc* hypotheses.

179 The real starting point for evolutionary genetic studies was the discovery
180 of a mutation of that linear structure. Inversions, which consist of replacing
181 a chromosome segment by its reverse order, were hypothesized by Sturtevant
182 (1921), based on the observation of differences in the arrangements of
183 five "corresponding"⁶ markers of chromosome 3 of *Drosophila simulans* and
184 *Drosophila melanogaster*. The inversion hypothesis was confirmed by adding
185 markers, while the comparative mapping of Sturtevant and Plunkett (1926),
186 as depicted in Figure 2, presents a striking visual argument for it⁷.

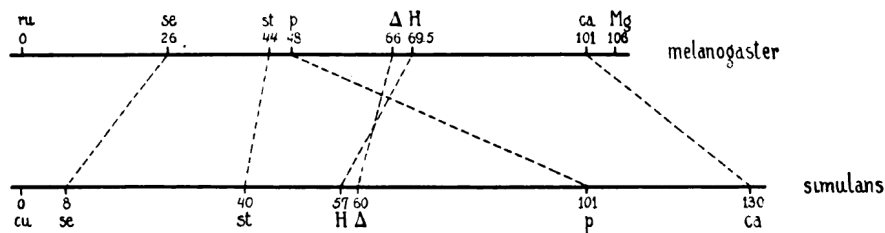


Fig. 2 Reproduced from Figure 1 of Sturtevant and Plunkett (1926): a visual argument for the existence of inversions and their utility for taxonomy. Linkage group 3 of *Drosophila simulans* (lower solid line) and *Drosophila melanogaster* (upper solid line) are compared. They consist in placing genes (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).

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

191 From this possibility of detecting a mutation by comparing genetic maps,
192 Sturtevant developed a comprehensive research project in line with his collab-
193 oration with Metz. It consisted in mapping the genes of different *Drosophila*
194 species, assessing the homologies between these genes and, from the chro-
195 mosome structure, reconstructing the evolution of the associated organisms
196 (Kohler, 1994). This research plan was not fully realized, although several
197 publications and many unpublished results⁸ confirm decisive advances and
198 reveal some challenges.

199 2.2 The statement of a mathematical problem

200 Among the challenges was the detection of several successive overlapping in-
201 versions. Comparing two arrangements differing by one inversion was easy.
202 However if several overlapping inversions may have occurred, which is likely
203 if more distant species are compared, an additional difficulty arose. In 1937,
204 Sturtevant, published with C. C. Tan, a Ph.D. student of himself and Dobzhans-
205 sky, a comparison of the arrangements of 38 genes along all chromosomes of
206 *Drosophila melanogaster* and *Drosophila pseudoobscura* (Sturtevant and Tan,
207 1937). The comparative maps, inferred from the homologies of genes deduced
208 from similar phenotypic effects, are reproduced in Figure 3. Inversions are
209 not as visible as in Figure 2 because the species are more distant, thus the
210 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 necessary⁹ 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 III can be transformed into alphabetical order by two inversions
as follows:

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

211 The first inversion concerns the segment *DEFA*, and the second the segment
212 *FEDCB*. The problem of computation of this number becomes tricky when
213 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)
 III 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*.

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

LHFEBADCKIJGM
 → ABEFHLDCCKIJGM
 → ABCDLHFEKIJGM
 → ABCDEFHLKIJGM
 → ABCDEFHGJIKLM
 → ABCDEFGHJIKLM
 → ABCDEFGHIJKLM

221 is one of the several possible *bona fide* sequences of six successive inversions¹⁰.
222 The "detailed analysis" called for by Sturtevant and Tan (see Figure 3) would
223 have to wait several decades before it became possible with the help of new
224 mathematical and computational techniques (Fertin et al., 2009). Besides its
225 anecdotal value however, this mention that the result was unsure is impor-
226 tant, because later on it was turned into an error: meaning that the modest
227 statement of ignorance was forgotten, while the number seven was taken for
228 granted. Together with two other approximations concerning the statistical
229 statement, this would lead to erroneously changing the biological conclusion.

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

249 2.3 The call for a professional mathematician

250 Going back over the statements of Sturtevant and Tan (those reproduced in
251 Figure 3), Sturtevant and Novitski announced that they had solicited the help
252 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 values¹². 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 0.85¹³.

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.

309 It is retrospectively vertiginous that Sturtevant and Novitski (1941), as-
310 sisted by a mathematician, claimed in 1941 to correct the intuitive statement
311 of Sturtevant and Tan (1937) in 1937¹⁴, but did not correct the real error that
312 was present in the 1937 paper (the number of inversions necessary to sort the
313 letters in alphabetical order). Instead they corrected, in the wrong direction,
314 the statement that the permutations were different from random ones, which
315 introduced two errors while failing to correct the only real one. The help of a
316 mathematician, which ordinarily would have been considered a good idea for
317 such a problem, was disastrous in this case, as it undermined, for the wrong
318 reasons, the sound intuitions that the biologists first had.

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

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

342 2.5 Epilogue: Theodosius Dobzhansky and Polytene Chromosomes

343 Of course the evolutionary genetics project Metz and Sturtevant had in 1914
344 has not stopped because of a few mathematical mistakes made in the 1930s.
345 One important reason is that it necessitated a prohibitive amount of work
346 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.

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

352 In 1933, Theophilus Painter invented an alternative technique (Painter,
353 1933, 1934) to detect similarities and differences between chromosome or-
354 ganizations. It consisted in hybridizing polytene chromosomes from different
355 strains or species of insects and visualizing chromosome inversions (Figure 5).
356 Immediately, geneticists and cytologists exploited the potential of this tech-
357 nique from a comparative and evolutionary perspective (Gannett and Greise-
358 mer, 2004). It has remained a widely used technique for studying rearrange-
359 ments until today.

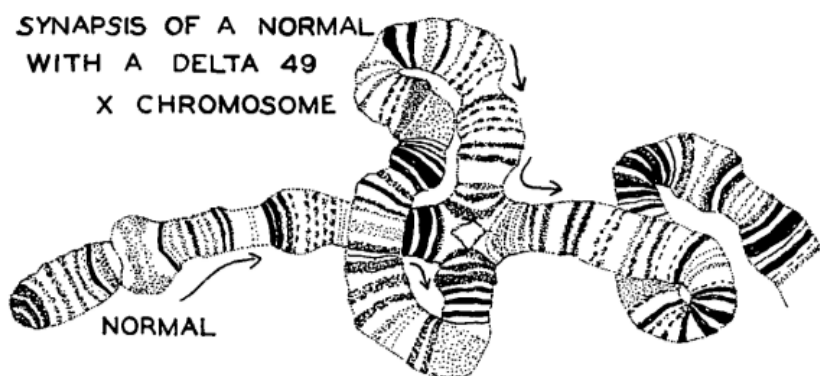


Fig. 5 Concatenated polytene X-chromosomes of *Drosophila melanogaster*, from Figure 1 in Painter (1934). An inversion on the "delta 49" variant strain is made visible by a loop on the concatenated giant chromosome. Reproduced with the kind permission of the *American Genetics Association* and the *Journal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

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

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

373 Dobzhansky published copiously and was expansive, extravagant, irreverent,
 374 religious and sensitive to celebrity: in short, the opposite of Sturtevant's per-
 375 sonality (Novitski, 2005).

376 Initially their differences yielded very productive and innovative scientific
 377 research. Dobzhansky collected many *pseudoobscura* strains from all around
 378 California and the United States¹⁵ and promoted the cytological studies, while
 379 Sturtevant took what was interesting for genetics and, in particular, his now
 380 old project of comparing chromosomes from an evolutionary perspective. The
 381 field strains collected by Dobzhansky and the Painter technique seemed ideal
 382 for studying the evolutionary history from chromosome inversions. Indeed,
 383 the two identified races of *Drosophila pseudoobscura* (see Figure 6) showed
 384 some intraracial variability. It was possible to classify all the strains according
 385 to the structure of their chromosomes, such that a hypothetical evolutionary
 386 history could be deduced from this classification. In 1936, the first phylogenetic
 387 tree whose branches are chromosome inversions was published (Sturtevant and
 388 Dobzhansky, 1936) (see Figure 6).

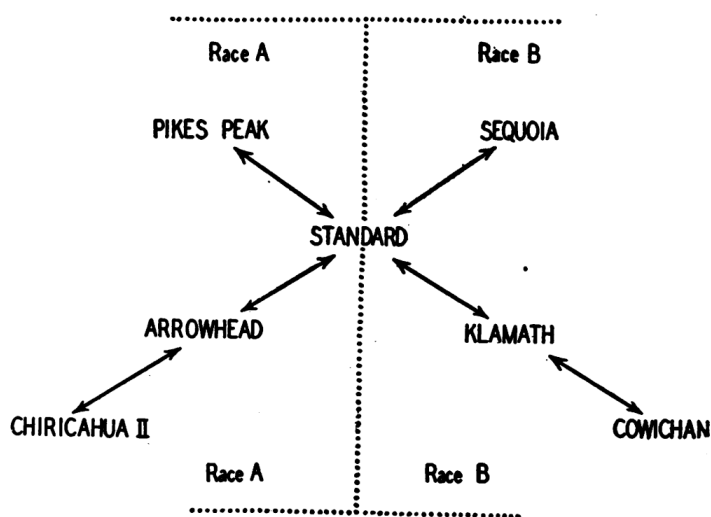


FIGURE 1

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.

389 An augmented phylogeny was published in 1938 by Dobzhansky and Sturte-
 390 vant, including 17 *pseudoobscura* strains (Dobzhansky and Sturtevant, 1938),
 391 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*.

392 authors had turned into opposition. The retirement of Morgan, aged 70 in
393 1936, and the inevitable succession, revealed ambitions, frustrations and fears
394 (Kohler, 1994). It prompted the beginning of the rift. The 1938 paper, the
395 last of their fruitful collaboration, was mainly Dobzhansky's work according
396 to himself (in his oral history interviews of 1962, cited by Kohler (1994)),
397 a situation which continued in the series "Genetics of Natural Populations"
398 (Dobzhansky and Queal, 1938). In 1940, Dobzhansky left Caltech for Col-
399 umbia¹⁶.

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

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

415 It is noticeable that, when the comparison of protein sequences was made
416 possible by new molecular techniques (Zuckerandl and Pauling, 1965), Dob-
417 zhansky did not engage in this field. Instead, he attacked this new trend in
418 molecular studies by taking a "naturalist's" point of view (Dobzhansky, 1963,
419 1966; Dietrich, 1998). This is symptomatic of how genetics in the first half of
420 the twentieth century on one side, and molecular evolution in the second half
421 on the other, have negotiated their place in evolutionary biology (Hagen, 1999;
422 Dietrich, 1998; Sommer, 2008). Dobzhansky, who could accommodate genetics
423 as a component of the study of evolution and participate in the unification
424 of genetics and evolution, could not accept the imperialistic statements of
425 Zuckerandl on protein comparisons, claiming that proteins were possibly the
426 only reliable materials with which to study evolution.

427 3 Discussion

428 3.1 Counting mutations as a computational biology problem

429 The introduction of measures, statistics and mathematics into evolutionary
430 studies and phylogeny has traversed all the twentieth century (Hagen, 2003;
431 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).

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

436 Computing evolutionary distances has been an important activity for es-
437 tablishing phylogenetic relationships. In the first half of the twentieth cen-
438 tury it was done for example with serological and immunological reactions
439 (de Chadarevian, 1996; Strasser, 2010b; Hagen, 2010), DNA hybridization
440 (Suárez-Díaz, 2014), which results hopefully would reflect the amount of di-
441 vergence between proteins or chromosomes.

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

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

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

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

477 Friden calculators from the biology department at Caltech, because he mis-
478 trusted all quantitative and automatic results¹⁷.

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

486 3.2 Interdisciplinary studies

487 Interdisciplinarity is also an element of this history, with ambiguous outcomes.
488 It was already not rare that mathematicians got involved in biological prob-
489 lems, or that biologists use statistics, or that biologists and mathematicians
490 would collaborate (Hagen, 2003). Sturtevant was able to use basic mathemat-
491 ics, had a "mathematical mind" according to his student Novitski (Novitski,
492 2005), and began as an autonomous and modest mathematician the compar-
493 ison of chromosomes (Sturtevant and Tan, 1937). Along the lines of this article,
494 a mathematical and a computational problem are modeled, and some solutions
495 arise. It is acknowledged that some results are "not yet certain". This might
496 have decided Sturtevant to contact Morgan Ward, from the mathematics de-
497 partment of his university. Sturtevant was able to model and solve to a certain
498 extent the mathematical problem, but was not "yet" certain of the result. He
499 already had in mind to check it and may be he had in mind to have it checked
500 by a professional. This interdisciplinary practice is not the kind practiced by
501 Dobzhansky and Wright at the same time, where both equally involved, com-
502 plementary. Ward is not involved in the general project, as witnessed by an
503 acknowledgement but no co-authorship. From this collaboration with a mathe-
504 matician, Sturtevant has made some advances on his problem, and also adopts
505 a less modest style concerning the results. He trusts in Ward more than in him-
506 self concerning mathematical results. Part of this rigorous and modest mind
507 has led him to engage in an erroneous direction.

508 3.3 The importance computational complexity

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

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

516 *pseudo-obscura* have not been reproduced for other species, nor the mathemat-
517 ical techniques have been refined, and the errors corrected by further work.

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

525 Thus one of the reasons why he did not achieve it at that time might
526 be because the mathematics raised by successive overlapping inversions is in-
527 tractable¹⁸. Not only are these problems hard¹⁹, but the possibility of their
528 resolution required the development of theoretical computer science and algo-
529 rithmics, whose starting point can be seen as concomitant with Sturtevant's
530 research. Although little known and often despised by most biologists, compu-
531 tational complexity is a constraint that can have an influence on the directions
532 taken by biological researches, and the history narrated here could be an ex-
533 ample of the influence of such a constraint. Computational complexity might
534 also explain why until today some biological processes are intensively studied
535 (point mutations in DNA) while others are much less quantified (repetitions,
536 inversions).

537 3.4 The importance of errors

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

544 Errors are nonetheless a good way to approach the ambiguous relationship
545 that scientists have to the truth, and can be characteristic of a knowledge
546 state, of a sociological or psychological context. Here, the errors point at some
547 challenges that mathematical and computational biology have to face regard-
548 ing computational complexity, interdisciplinary studies, and how a number,
549 associated with a caution note, can be taken four years later for a trustworthy
550 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).

551 **Acknowledgements**

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

559 **References**

- 560 Anderson, E. 1937, Jul. Cytology in its relation to taxonomy. *The Botanical*
561 *Review* 3(7), 335–350.
- 562 Babcock, E. B. and G. L. Stebbins 1938. *The American species of Crepis: Their*
563 *interrelationships and distribution as affected by polyploidy and apomixis.*
564 Washington, D.C: Carnegie Institution of Washington.
- 565 Bafna, V. and P. A. Pevzner 1996, February. Genome rearrangements and
566 sorting by reversals. *SIAM J. Comput.* 25(2), 272–289.
- 567 Belevin, P. 2007. *Seeking Wisdom: From Darwin to Munger.* PCA Publica-
568 tions.
- 569 Bosch, G. 2018, February. Train PhD students to be thinkers not just special-
570 ists. *Nature* 554(7692), 277–277.
- 571 Bowler, P. J. 2003, 07. *Evolution.* University of California Press.
- 572 Boyden, A. 1934. Precipitins and phylogeny in animals. *The American Natu-*
573 *ralist* 68(719), 516–536.
- 574 Brehm, A. 1990. *Phylogénie de neuf espèces de Drosophila du groupe obscura*
575 *d'après les homologues de segments des chromosomes polytènes.* Ph. D. the-
576 sis, Université de Lyon 1.
- 577 Bridges, C. B. 1917. Deficiency. *Genetics* 2, 445–465.
- 578 Camin, J. H. and R. R. Sokal 1965, sep. A method for deducing branching
579 sequences in phylogeny. *Evolution* 19(3), 311–326.
- 580 Caprara, A. and G. Lancia 2000. Experimental and statistical analysis of
581 sorting by reversals. In D. Sankoff and J. H. Nadeau (Eds.), *Comparative*
582 *Genomics*, pp. 171–183. Springer.
- 583 Carson, H. L. and K. Y. Kaneshiro 1976. Drosophila of hawaii: systematics
584 and ecological genetics. *Annual Review of Ecology and Systematics* 7(1),
585 311–345.
- 586 Castle, W. E. 1918, Feb. Is the arrangement of the genes in the chromosome
587 linear? *Proc Natl Acad Sci U S A* 5(2), 25–32.
- 588 ——— 1919, Nov. Are genes linear or non-linear in arrangement? *Proc Natl*
589 *Acad Sci U S A* 5(11), 500–506.
- 590 Darden, L. 2005, jun. Relations among fields: Mendelian, cytological and
591 molecular mechanisms. *Studies in History and Philosophy of Science Part*

- 592 *C: Studies in History and Philosophy of Biological and Biomedical Sci-*
593 *ences* 36(2), 349–371.
- 594 de Chadarevian, S. 1996, Sep. Sequences, conformation, information: Bio-
595 chemists and molecular biologists in the 1950s. *Journal of the History of*
596 *Biology* 29(3), 361–386.
- 597 de Chadarevian, S. and H. Kamminga (Eds.) 1998. *Molecularizing Biology and*
598 *Medicine New Practices and Alliances, 1920s to 1970s*. Taylor and Francis.
- 599 Dietrich, M. R. 1994, Mar. The origins of the neutral theory of molecular
600 evolution. *Journal of the History of Biology* 27(1), 21–59.
- 601 ——— 1998. Paradox and persuasion: negotiating the place of molecular evolu-
602 tion within evolutionary biology. *J Hist Biol* 31(1), 85–111.
- 603 Dietrich, M. R. 2016. History of molecular evolution. In *Encyclopedia of*
604 *Evolutionary Biology*. Elsevier.
- 605 Dobzhansky, T. 1930. Translocations involving the third and the fourth chro-
606 mosomes of drosophila melanogaster. *Genetics* 15(4), 347–399.
- 607 ——— 1963. Evolutionary and population genetics. *Science* 142(3596), 1131–
608 1135.
- 609 ——— 1966. Are naturalists old-fashioned? *The American Naturalist* 100(915),
610 541–550.
- 611 Dobzhansky, T. and J. R. Powell 1975. *Drosophila pseudoobscura* and its
612 american relatives, *drosophila persimilis* and *drosophila miranda*. In R. King
613 (Ed.), *Invertebrates of Genetic Interest*, pp. 537–587. Plenum Press.
- 614 Dobzhansky, T. and M. Queal 1938. Genetics of natural populations. i. chro-
615 mosome variation in populations of drosophila pseudoobscura inhabiting
616 isolated mountain ranges. *Genetics* 23(2), 239.
- 617 Dobzhansky, T. and A. H. Sturtevant 1938, Jan. Inversions in the chromosomes
618 of drosophila pseudoobscura. *Genetics* 23(1), 28–64.
- 619 Dutrillaux, A.-M. and B. Dutrillaux 2012. Chromosome analysis of 82 species
620 of scarabaeoidea (coleoptera), with special focus on nor localization. *Cyto-*
621 *genetic and genome research* 136, 208–219.
- 622 Fertin, G., A. Labarre, I. Rusu, E. Tannier, and S. Vialette 2009. *Combina-*
623 *torics of Genome Rearrangements*. London: MIT press.
- 624 Firestein, S. 2015. *Failure*. oxford university press.
- 625 Galvão, G. R. and Z. Dias 2015, January. An audit tool for genome rearrange-
626 ment algorithms. *J. Exp. Algorithmics* 19, 1.7:1.1–1.7:1.34.
- 627 Gannett, L. and J. R. Greisemer 2004. Classical genetics and the geography
628 of genes. In Rheinberger and Gaudilliere (Eds.), *Classical Genetic Research*
629 *and Its Legacy*, pp. 57–88. London and New York: Routledge.
- 630 Hagen, J. B. 1982. *Experimental Taxonomy, 1930-1950: The Impact of Cy-*
631 *tology, Ecology, and Genetics on Ideas of Biological Classification*. Ph. D.
632 thesis, Oregon State University.
- 633 Hagen, J. B. 1984, Jun. Experimentalists and naturalists in twentieth-century
634 botany: Experimental taxonomy, 1920–1950. *Journal of the History of Bi-*
635 *ology* 17(2), 249–270.
- 636 ——— 1999. Naturalists, molecular biologists, and the challenges of molecular
637 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-

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