LACK OF CONSTANT GENOME ENDS IN *STREPTOMYCES COELICOLOR*

D. A. HOPWOOD

Department of Genetics, University of Glasgow, Glasgow, Scotland

Received June 13, 1966

IT was recently shown (HOPWOOD 1965b) that all known markers of *Streptomyces coelicolor* are linked on a circular map. The question then arises: what is the topology of the genome? We may consider three possibilities: (1) The genome is a closed loop; (2) The genome is open, but its ends vary in position, so that individual chromosomes are circular permutations of one another (STREISINGER, EDGAR and DENHARDT 1964); (3) The genome is open, with ends at constant positions. In the absence of special assumptions (1) and (2) would lead to map circularity, as in *Escherichia coli* and bacteriophage T4 respectively. (3) would also do so if the products of even numbers of crossovers were recovered, but not those of odd numbers (STAHL and STEINBERG 1964). Now, there is evidence (HOPWOOD, manuscript in preparation) that the zygotes of *S. coelicolor* are partial diploids, consisting of a complete chromosome from one parent and an incomplete one from the other. Such merozygotes, as in other bacterial species, give rise to recombinant progeny with complete haploid genomes only by even numbers of crossovers; therefore when a sample of haploid recombinants from a cross is analysed, map circularity results, and would do so even if the genome were open, with ends in constant positions. This situation, then, prevents us from deducing lack of constant genome ends from the finding of map circularity.

The present paper describes an analysis of individual *heteroclones* (SERMONTI, MANCINELLI and SPADA-SERMONTI 1960), which are heterozygous colonies originating from spores containing partial diploid nuclei (HOPWOOD, SERMONTI and SPADA-SERMONTI 1963). It is probable that each of the nuclei is derived, by an odd number of crossovers, from a merozygote containing a chromosomal fragment of a certain length, and that the length of the diploid region in the heteroclone corresponds to the length of this fragment (HOPWOOD, in preparation), but this is not a necessary part of the argument in the present paper. Heteroclones are obtained by plating the products of a cross on a medium selective for wild-type alleles, one from each parent, at two closely linked nutritional loci. The nuclei of the heteroclones are diploid at these closely linked loci, and the diploid region extends, for distances which vary from one heteroclone to another, on either side of these loci. Markers within the diploid region are (with rare exceptions) heterozygous, and those outside it are hemizygous, so that the length of the diploid region in each heteroclone can be determined by analysing the genotypes of the haploid segregants to which the heteroclone gives rise. The present

study shows that the diploid segments of different heteroclones overlap to cover the whole of the circular map, and indicates that the genome either has ends which vary in position, or has no ends.

MATERIALS AND METHODS

Two strains bearing the markers *argA1 hisC9 strA1* and *hisA1 cysA15 nicA3 uraA1 tps30* (locus abbreviations are explained in the caption to Figure 1) were crossed on complete medium (HOPWOOD and SERMONTI 1962) and spores were plated on minimal medium lacking histidine, but containing arginine, cystine, nicotinamide and uracil. (The first strain also bore the mutants *pheA1* (phenylalanine) and *proA1* (proline) but these were ignored by adding phenylalanine

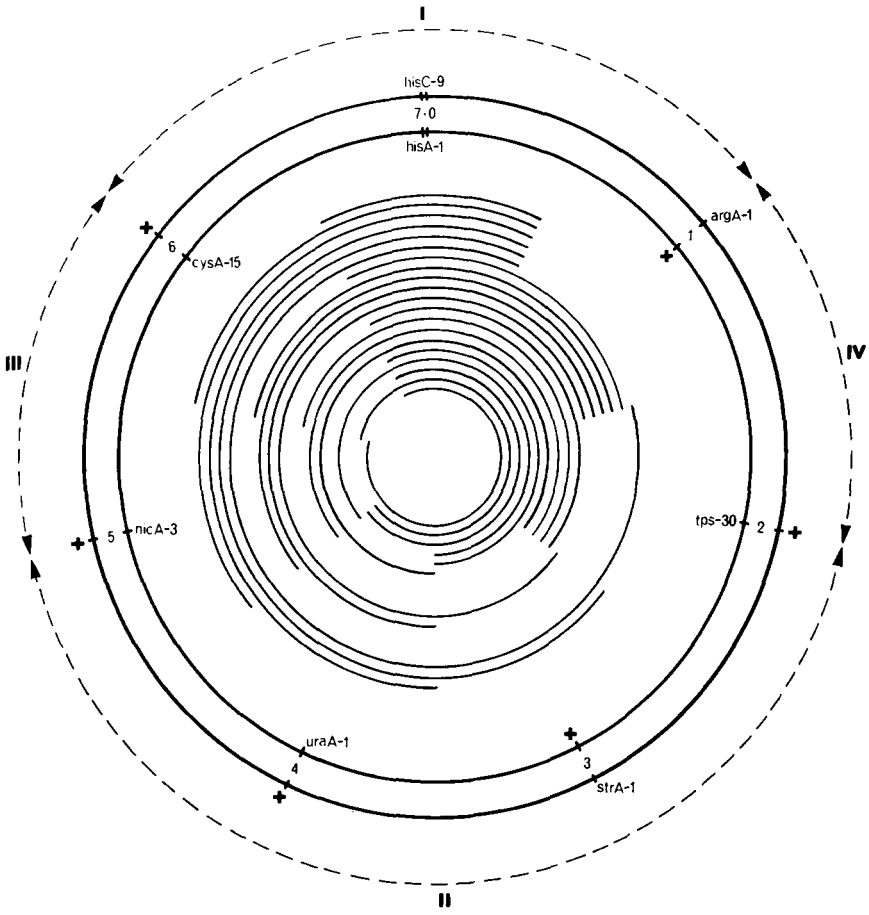

FIGURE 1.—Representation of the cross and its results. The markers of the two parents are on the two outer unbroken circles. Loci are arbitrarily spaced equidistantly, and are numbered clockwise from 0 to 7. Locus symbols: *arg, cys, his, nic, ura,* requirement for arginine, cyst(e)ine, histidine, nicotinamide, uracil respectively; *str,* resistance to streptomycin; *tps,* temperature-sensitive. Segments of the outermost, broken circle: I and II are the previously identified linkage groups; III and IV indicate the regions of the circular map connecting these two groups. The diploid regions in 20 of the 24 classes of heteroclones in Table 1 (classes *w, x, y, z* omitted) on the hypothesis of no constant genome ends are represented by arcs in the central part of the figure.

and proline to all media). The colonies that developed on the selective medium were therefore either haploid recombinants between the two closely linked histidine loci, *hisA1* and *hisC9*, which are less than one map unit apart, or heteroclones diploid for these loci. The heteroclones were distinguished from the haploids by replica plating (SERMONTI *et al* 1960).

In previous studies (SERMONTI *et al.* 1960; HOPWOOD and SERMONTI 1962; HOPWOOD *et al.* 1963; SERMONTI and HOPWOOD 1964; HOPWOOD 1965a) each heteroclone was analysed by determining the full genotypes of a sample of 50 or more haploid segregants from it. In the present experiment, in order to study a large sample of heteroclones, each heteroclone was analysed merely for heterozygosity or otherwise of its markers, as follows. Master plates of minimal medium supplemented with all the nutritional requirements of the parent strains were marked out into eight sectors, in each of which a spore suspension from a single heteroclone was spread, at a dilution giving between 50 and a few hundred colonies per streak. After a few days incubation at 30°C, the master plates were replicated to diagnostic media. For each parental marker it was determined whether a streak derived from a particular heteroclone consisted of: (a) all wild-type colonies; (b) all mutant colonies; (c) a mixture of wild-type and mutant. (a) or (b) indicated hemizygosity (or rarely homozygosity) of the marker, while (c) indicated heterozygosity. The distinction of (c) from (a) was facilitated when the mutant was leaky, so that two kinds of colony, normal and slow-growing, were seen on the diagnostic plate in case (c). *cysA15* and *nicA3* were sufficiently leaky for easy scoring on plates lacking cystine and nicotinamide respectively, while *argA1* and *uraA1* were scored on plates supplemented with one-tenth the normal concentration of the relevant growth factor. *tps* 30 (temperature-sensitive) was sufficiently leaky for the recognition of mixed streaks at 38° (HOPWOOD 1966). *strA1* was scored by comparing the number of colonies growing on replica plates with and without streptomycin.

RESULTS AND DISCUSSION

Lacks of constant genome ends: Figure 1 is a representation of the cross, in which regions I and II are the two linkage groups previously identified, and III and IV are the regions of the circular map joining groups I and II.

The results of the cross are analysed in the light of the following facts: (1) the genome is continuous, at least in some heteroclones, over region I, since linkages within this region have been established or confirmed by the analysis of single heteroclones (HOPWOOD and SERMONTI 1962; HOPWOOD 1965a, 1966), and the same is true of region II; (2) the genome is continuous over at least one of regions III and IV, since there is linkage between markers in regions I and II (HOPWOOD 1965b). It follows from these considerations that, if the genome has constant ends, the ends must lie either in III or in IV. We therefore choose among three hypotheses about genome ends: ends in III; ends in IV; or no constant ends.

The 447 heteroclones tested fall into 24 classes (Table 1), on the basis of heterozygosity or nonheterozygosity of the various markers. Three hundred and eighty-three heteroclones have one or more markers not heterozygous, and these allow the limits of the diploid segment to be determined. The remaining class, *z*, consisting of 64 heteroclones, represents a heterogeneous group. Some members of this class are heteroclones in which both ends of the diploid segment fall in the same one of the seven map intervals (or conceivably heteroclones in which the diploid segment lacks ends). However, class *z* also contains an unknown number of heteroclones with shorter diploid segments, but with some contamination by parental spores picked up from the plates of selective medium during the isolation of the heteroclones, resulting in apparent heterozygosity for all markers. Such

TABLE 1

Patterns of heterozygosity in a sample of heteroclones from the cross represented in Figure 1

Class of heteroclones*	Loci at which heteroclones are heterozygous (+) or not heterozygous (·)							Observed number of heteroclones
	ura (4)	*nic* (5)	*cys* (6)	*his* (7, 0)	*arg* (1)	*tps* (2)	*str* (3)	
07	·	·	·	+	·	·	·	48
06	·	·	+	+	·	·	·	12
05	·	+	+	+	·	·	·	12
04	+	+	+	+	·	·	·	8
03	+	+	+	+	·	·	+	1
02	+	+	+	+	·	+	+	3
17	·	·	·	+	+	·	·	132
16	·	·	+	+	+	·	·	49
15	·	+	+	+	+	·	·	60
14	+	+	+	+	+	·	·	12
13	+	+	+	+	+	·	+	3
27	·	·	·	+	+	+	·	14
26	·	·	+	+	+	+	·	5
25	·	+	+	+	+	+	·	6
24	+	+	+	+	+	+	·	6
37	·	·	·	+	+	+	+	1
36	·	·	+	+	+	+	+	0
35	·	+	+	+	+	+	+	2
47	+	·	·	+	+	+	+	1
46	+	·	+	+	+	+	+	1
57	+	+	·	+	+	+	+	3
w	·	+	·	+	+	·	·	2
x	+	·	·	+	+	+	·	1
y	·	+	·	+	+	+	·	1
z	+	+	+	+	+	+	+	64
Total								447

* Numbered, where appropriate, according to the terminal heterozygous loci.

parental contamination was minimised by isolating the heteroclones from selective plates prepared from highly diluted spore suspensions, but was not eliminated entirely. Fortunately the *z* class can be ignored in the present discussion, and also in the statistical analysis of the data (EDWARDS 1966).

It is evident that, on any of the three hypotheses, each of the heteroclones usually contains a single uninterrupted diploid segment. This is in agreement with a hypothesis on the nature of heteroclones (HOPWOOD, in preparation) which implies that only a single diploid region can occur in each. There are, however, certain heteroclones, which we shall call "aberrant", in which heterozygous markers are separated from each other by markers that are not heterozygous. Occasional heteroclones of this type may be due to homozygosity of one or more markers within the diploid region (HOPWOOD and SERMONTI 1962). In choosing

among the three hypotheses, "aberrant" heteroclones have to be taken into account.

On the hypothesis that the genome does not have constant ends, that is that it may be continuous, at least in some heteroclones, over any part of the circular map, the diploid segment in 379 of the 383 heteroclones (ignoring class *z*) can be represented by an unbroken arc as in Figure 1, leaving only four "aberrant" heteroclones (classes *w*, *x*, *y*). On the two alternative hypotheses, the number of "aberrant" heteroclones is much higher: 115 (classes *05*, *04*, *03*, *02*, *15*, *14*, *13*, *25*, *24*, *35*, *w*, *x*, *y*) if the genome ends are in region III, and 43 (classes *27*, *26*, *25*, *24*, *37*, *36*, *35*, *47*, *46*, *57*, *w*, *x*, *y*) if the ends are in IV.

There is no compelling *a priori* argument against the existence of large numbers of "aberrant" heteroclones. However, good grounds for rejecting the two hypotheses of constant ends emerge when we analyse their implications for the various classes of heteroclones that would be "aberrant" on these hypotheses. All heteroclones are selected to be heterozygous at *his*; if we consider only those heteroclones that are *not* heterozygous at *str*, the locus farthest from *his* on the circular map (this entails the exclusion of only 15 of the 383 heteroclones under consideration), the map is divided by *his* and *str* into two arcs, which should be independent of each other in respect of heterozygosity of markers. On the other hand, markers on the same arc should be independent of each other in respect of heterozygosity only if the markers are separated by genome ends; in that case, *all* heteroclones heterozygous for the marker farther from *his* on that arc would be "aberrant".

Heterozygosity of the four markers flanking regions III and IV is analysed in Table 2. As expected, there is independence between markers on opposite arcs: *nic* and *cys* on one; *arg* and *tps* on the other. In contrast, markers on the same arc show a highly significant lack of independence. In this sample *tps* is never heterozygous unless *arg* is also heterozygous, indicating genome continuity over interval IV, while *nic* is very rarely heterozygous when *cys* is not (the three exceptions,

TABLE 2

Numbers of heteroclones heterozygous at various pairs of loci: 2 × 2 tests of independence

		cys		*arg*		*nic*	
		Hetero-zygous	Not hetero-zygous	Hetero-zygous	Not hetero-zygous	Hetero-zygous	Not hetero-zygous
tps	heterozygous	17	16	33	0	13	20
	not heterozygous	153	182	255	80	94	241
	probability	>0.05		0.0002		>0.05	
cys	heterozygous	. .	. .	138	32	104	66
	not heterozygous	. .	. .	150	48	3	195
	probability			>0.05		$\ll 10^{-6}$	
arg	heterozygous	. .	. .	. .	. .	87	201
	not heterozygous	. .	. .	. .	. .	20	60
	probability					>0.05	

classes *w* and *y*, are "aberrant" on all three hypotheses), indicating genome continuity over interval III.

We therefore conclude that the genome lacks constant ends. The diploid segments in different heteroclones overlap to cover, among them, the whole of the circular map (Figure 1).

Quantitative evaluation of the data: Further information can be obtained by a statistical analysis of the data in Table 1, for which I am greatly indebted to DR. A. W. F. EDWARDS. In the absence of experimental evidence as to how the chromosomal fragments which give rise to the diploid regions of heteroclones arise, this analysis is based on formal models. In the accompanying paper, EDWARDS (1966) considers two of these. In both, a complete chromosome is fragmented by random breaks to give a population of fragments, from which the experiment has selected those bearing the *his* loci. On hypothesis (1), the complete chromosome is assumed to be a closed loop (circle), while on hypothesis (2), complete chromosomes are assumed to be open circular permutations, each notionally derived from a circle by a random mandatory break.

It turns out that the numerical consequences of the two hypotheses are closely similar, as reflected in similar estimates of the seven map distances (Table 3). The agreement of the observed numbers of heteroclones in each class with those expected on the two hypotheses is almost equally good (Table 4). This prevents us from excluding either hypothesis, but gives confidence in the internal consistency of the data.

A third hypothesis, borrowed from *Escherichia coli*, in which a circle is opened by a random mandatory break to give an open chromosome, which is then extruded from one parent to the other, with further random breakage during extrusion, turns out to be more complex mathematically and it has not so far been possible to investigate its consequences. However, it is unlikely that they would differ enough from those of the other hypotheses for a choice to be made between them.

The estimates of the map distances derived from the present data fill an important gap in our knowledge of the linkage map of *S. coelicolor*. From *cys* to *arg* and from *tps* to *ura*, in a clockwise direction, the map is abundantly marked

TABLE 3

Percentage map distances on two hypotheses: (1) no mandatory break; (2) one mandatory break

Interval	Percentage map distance	
	Hypothesis (1)	Hypothesis (2)
his–arg	3.82	3.62
arg–tps	30.21	28.53
tps–str	16.74	18.28
str–ura	14.23	15.63
ura–nic	16.60	16.40
nic–cys	7.12	6.86
cys–his	11.27	10.68
Mean number of (nonmandatory) breaks:	6.41	5.83

TABLE 4

Goodness-of-fit of observed numbers of heteroclones in each class with those expected on two hypotheses: (1) no mandatory break; (2) one mandatory break

Class of heteroclones	Expected number on hypothesis (1)	Observed number	Expected number on hypothesis (2)
07	43.33	48	43.12
06	14.98	12	15.27
05	16.95	12	17.43
04	5.34	8	5.53
03	2.36	1	2.00
02	1.05	3	0.61
17	133.83	132	134.33
16	46.26	49	47.21
15	52.37	60	53.21
14	16.50	12	16.52
13	7.28	3	5.77
27	14.83	14	16.01
26	5.13	5	5.49
25	5.80	6	5.97
24	1.83	6	1.72
37	4.61	1	4.06
36	1.59	0	1.35
35	1.80	2	1.41
47	2.02	1	1.31
46	0.70	1	0.42
57	0.39	3	0.21
Total	378.95	379	378.95
*χ^2_6	8.93		9.98
P	0.18		0.13

* Small classes combined as indicated by brackets.

(HOPWOOD 1966), so that satisfactory estimates of map distances were obtained previously, but estimates for the intervening regions were poor or lacking owing to a dearth of available markers. Both of these intervals turn out to be about a quarter of the total map. The significance of this finding is discussed elsewhere (HOPWOOD 1966).

The present study gives no measure of the length of the linkage map in crossover units. However, we can obtain an estimate in these terms by equating the percentage map distances in Table 3 with map distances in centimorgans over those intervals for which reasonable estimates are available (Table 5). We see that there is good agreement between the relative lengths of the four intervals in terms of the probability of breakage of the chromosome to produce the heteroclone fragment, and in terms of crossover units. Equating the two, the total length of the genome turns out to be about 260 centimorgans.

Conclusion: The experiment reported here does not provide a definitive answer

TABLE 5

Comparison of percentage map distances on the two hypotheses with map distances in centimorgans

Interval	Percentage map distance		Map distance in cM*	
	Hypothesis (1)	Hypothesis (2)		
his–arg	3.82 (8)	3.62 (8)	12	(10)
tps-str	16.74 (36)	18.28 (38)	49	(40)
str–ura	14.23 (31)	15.63 (32)	32	(26)
cys–his	11.27 (25)	10.68 (22)	30	(24)
	46.06 (100)	48.01 (100)	123	(100)
47% of the map = 123 cM: total map length = 262 cM				

* Calculated by HALDANE's (1919) formula (no interference) from recombination percentages in HOPWOOD and SERMONTI (1962) and in unpublished data.

to the question posed in the introduction: what is the topology of the genome of *S. coelicolor*? It does, however, exclude the possibility that the genome has ends which are invariably in the same positions. Absence of constant genome ends is thus a characteristic in which *S. coelicolor* resembles the eubacterium *Escherichia coli;* whether, like that of *E. coli* (CAIRNS 1963), the genome of *S. coelicolor* can actually be a closed loop, remains to be determined.

I am much indebted to DR. J. H. RENWICK for suggestions on the evaluation of the data reported in this paper.

SUMMARY

The genome of *Streptomyces coelicolor* is about 260 map units long and has no constant ends.

LITERATURE CITED

CAIRNS, J., 1963 The bacterial chromosome and its manner of replication as seen by autoradiography. J. Mol. Biol. **6:** 208–213.

EDWARDS, A. W. F., 1966 Linkage on a circular map. Genetics **54:** 1185–1187.

HALDANE, J. B. S., 1919 The combination of linkage values, and the calculation of distance between the loci of linked factors. J. Genet. **8:** 299–309.

HOPWOOD, D. A., 1965a New data on the linkage map of *Streptomyces coelicolor*. Genet. Res. **6:** 248–262. —— 1965b A circular linkage map in the Actinomycete *Streptomyces coelicolor*. J. Mol. Biol. **12:** 514–516. —— 1966 Nonrandom location of temperature-sensitive mutants on the linkage map of *Streptomyces coelicolor*. Genetics **54:** 1169–1176.

HOPWOOD, D. A., and G. SERMONTI, 1962 The genetics of *Streptomyces coelicolor*. Advan. Genet. **11:** 273–342.

HOPWOOD, D. A., G. SERMONTI, and I. SPADA-SERMONTI, 1963 Heterozygous clones in *Streptomyces coelicolor*. J. Gen. Microbiol. **30:** 249–260.

SERMONTI, G., and D. A. HOPWOOD, 1964 Genetic recombination in Streptomyces. pp. 223–251. *The Bacteria, Volume 5, Heredity*. Edited by I. C. GUNSALUS and R. Y. STANIER. Academic Press, New York.

SERMONTI, G., A. MANCINELLI, and I. SPADA-SERMONTI, 1960 Heterogeneous clones (heteroclones) in *Streptomyces coelicolor* A3(2). Genetics **45:** 669–672.

STAHL, F. W., and C. M. STEINBERG, 1964 The theory of formal phage genetics for circular maps. Genetics **50:** 531–538.

STREISINGER, G., R. S. EDGAR, and G. H. DENHARDT, 1964 Chromosome structure in phage T4. I. Circularity of the linkage map. Proc. Natl. Acad. Sci. U.S. **51:** 775–779.