arXiv:1708.05255v3 [math.CT] 6 Oct 2017

CATEGORY THEORY FOR GENETICS

RÉMY TUYÉRAS

Abstract. We introduce a categorical language in which it is possible to talk about DNA sequencing, alignment methods, CRISPR, homologous recombination, haplotypes, and genetic linkage. This language takes the form of a class of limit-sketches whose categories of models can model different concepts of Biology depending on what their categories of values are. We discuss examples of models in the category of sets and in the category of semimodules over the Boolean semi-ring $\{0, 1\}$. We identify a subclass of models in sets that models the genetic material of living beings and another subclass of models in semimodules that models haplotypes. We show how the two classes are related via a universal property.

1. Introduction

1.1. **Short presentation.** The goal of the present article is to define a type of algebraic structure in which it is possible to *do genetics*. Even though the proposed structures are completely algebraic, we will see that they also enable us to retrieve well-known statistical tools, such as the mapping functions used in genetic linkage [29, 26, 9], which map the recombination frequency of two DNA segments as a function of their distance on the chromosome and which usually take the form of cumulative distribution functions. The present paper should therefore be seen as an effort to clarify the tools of genetics through algebra and, more specifically, category theory, rather than a work that only restricts itself on describing the algebraic aspects of genetics.

1.2. **Motivations.** Our goal is to construct a bridge between two completely disconnected domains of science, specifically genetics and category theory. The former is well-known for its complexity while the latter is recognized for its clarity and expressive power [27, 4, 15].

The language of the present paper is rather mathematical, but the results and definitions that it contains always try to capture the biological reality. Note that some terms might be used in a biological sense while others might be used in a mathematical one – this will usually be specified. For instance, the sentence "a structure in which it is possible to do genetics" means that we want to define a formal language rather than a model of some particular living body. The need for such an abstraction, in Biology, has, for example, been recognized in [12].

In this respect, our structures should be seen as formal environments in which it would be possible to express a problem and say things about its solution, all within the same language. Note that it is very common, in Mathematics, to study the language in which it would be possible to solve a problem before even trying to solve the problem itself. The reason being that depending on the language one uses, the solution to the problem may or may not appear. For instance, the fact that a trait can be transmitted from grand-parents to grand-children without being expressed via the intermediate generation can be explained via Mendel's rules of inheritance, but not via Darwin's theory of natural selection.

In addition of offering a formalism, the present paper also tries to tackle technical and conceptual problems of various parts of genetics.

For example, phylogenetics have been recognized to need more clarity in order to be endowed with more satisfactory computerized procedures [18] and the set of possible operations

This research was supported by the AFOSR grant, *Categorical approach to agent interaction*, FA9550-14-1-0031 and the AFOSR grant, *Pixel matrices and other compositional analyses of interconnected systems*, FA9550-17-1-0058.

that can underlay alignment methods [25, Chap. 1] remains to be organized in the form of a theoretical framework, which no-one seems to have produced yet [19]. In this paper, we define a limit-sketch that comprises the most basic operations of alignment methods, such as insertion of gaps, cutting of DNA patches, concatenation of DNA patches and homologous recombination. More complex operations, such as duplications, transpositions, reversions, deletion, insertion and substitution mutations would then need to be expressed at the level of the models for this theory (see Examples 3.30; 3.32; 3.34; 4.68 & 4.69). This structural hierarchization, which pertains to the language of category theory, goes in direction of the program suggested in [19, 18] by trying to "[recognize] mechanisms rather than assuming that all the variation occurs at random" [19, page 156, right col., l. 5]. For instance, distinguishing homologous recombination from mutations by setting them at different levels (namely, that of the theory and that of the models, respectively) translates the fact that homologous recombination is more of a systematic event pertaining to the biological reality while mutations are more of a set of possible events pertaining to Evolution [20, Chap. 3].

Another example is in genetic linkage, where the current form of mapping functions do not quite fit recombination models, mainly because they do not manage to model cross-over interference [26, page 3, right col.]. It is also suggested [26, *Loc. cit.*] that the measuring of cross-over interference should be done between potentially-separated DNA patches rather than necessarily-adjacent ones. In this article, we will be able to specify such disconnected interference relations via the *cones* (see section 2.7) of our limit-sketch. In section 5, we will also see how these cones allow us to specify probability spaces of recombination events so that we will be able to recover Haldane's mapping functions [9]. We will conclude that the flexibility of our language makes it a good candidate for providing a framework in which it is possible to talk about multi-locus genetic linkage.

Finally, it is interesting to note that, both the hierachization of biological operations (such as the distinction between homologous recombination and mutations) and the specification of topologies (indicating where the recombination events occur) are shown to be parameters that can significantly determine the shape of evolutionary trees [2]. It is even concluded that methods being able to manage the space of evolutionary trees resulting from the space of these parameters is very much needed. We would here suggest that such a scheme would first require a formalization of the latter. In this article, we propose to formalize this space of parameters in terms of limit-sketches and models for these.

1.3. **Road map and results.** The goal of the paper is to define a class of theories, called *chromologies*, whose models, called *pedigrads*, can recover various aspects of genetics. We will start by defining chromologies in section 2, from section 2.1 to section 2.8 while the models (pedigrads) for these theories will be defined in sections 2.9 & 2.10. Intuitively, chromologoies will allow us to do all sorts of basic DNA manipulations such as DNA sequencing, alignment methods, CRISPR [23] and homologous recombination whereas the pedigrads will allow us to give a context to these operations (which can be handled differently depending on the environment in which they are processed).

In section 3, we define a class of canonical pedigrads taking values in the category of sets. The images of their underlying functors will be seen as sets containing DNA sequences. These canonical pedigrads will be used in section 4 to generate new pedigrads in a category of semimodules over a particular semi-ring. The examples and illustrations given in section 3 will be important to understand the content of section 4.

In section 4, we will construct a class of canonical pedigrads (see Definition 4.50 and Theorem 4.62) in the category of semimodules over the Boolean semi-ring $B_2 = \{0, 1\}$. The elements belonging to the images of their underlying functors can be seen as theoretical 'haplotypes'. Then, in sections 4.13 & 4.14, we will define a larger class of pedigrads , called *presentable pedigrads*, for which our class of canonical pedigrads will satisfy a universal property (see Theorem 4.72). Presentable pedigrads will be equipped with a type of morphism that can model polymorphic DNA mutations (see Example 4.68) as well as the usual transcription operations (see Example 4.69). We will see that the expressive power of

semimodules, and, more specifically, the equations that they satisfy, will allow us to express biological phenomena such as nullomers and other selective behaviors resulting from RNA translation (see Example 4.71).

Finally, in section 5, we will see how it is possible to recover the 'mapping functions' [9] expressing the genetic distance between two markers on a given chromosome. Section 5.2 will suggest various ways of refining the obtained mapping functions.

1.4. **Acknowledgments.** I would like to thank David Spivak and Eric Neumann for very useful discussions, remarks and questions regarding the content of this paper.

2. Chromologies and Pedigrads

The goal of this section is to introduce a set of theories whose models try to capture the logic of genetics. To justify why our theories look the way they do, we need to recall a few facts regarding the reconstruction of theories in general. First, recall that, classically, models for theories are defined as sets equipped with some operations. For instance, a 'ring' is a set R equipped with two operations $\cdot : R \times R \to R$ and $+ : R \times R \to R$ making certain diagrams commute.

Now, more categorically, rings are also product-preserving functors from a certain product sketch[1] `Ring` (the theory) to the category **Set** of sets and functions [5]. This functorial point of view was introduced by Lawvere [11] in 1963 via the concept of what is now called a *Lawvere theory* – the theory `Ring` being an example. The advantage of functors over sets equipped with functions is that functors allow us to clearly distinguish between what is intrinsically true in a model (via the theory) and what can occasionally be true in the model. Then, the formalism accompanying the language of functors allows us to more carefully think about the mechanisms governing the models.

Since Lawvere theories were meant to capture the logic of algebraic structures equipped with multivariate functions, their objects were taken to be the set of natural numbers in order to specify the arities of the functions. Along those lines, since the goal of the present section is to define a theory that captures the logic of genetics and whose operations takes DNA patches as inputs, the objects of our theory will look like DNA segments. Note that, while, in rings, one *adds* and *multiplies* terms together, in genetics, one *cuts*, *aligns* and *recombines* DNA strands together. Therefore, our theory will be based on these operations.

For illustration, an integer object in a Lawvere theory can easily be represented as a finite sequence of atoms; e.g the object 6 would be represented by six atoms as follows.

$$6 = (\bullet\bullet\bullet\bullet\bullet\bullet) \tag{2.1}$$

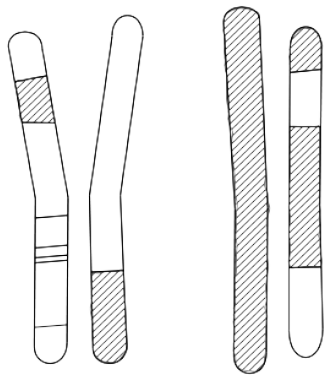
These atoms can make it easier to see how the functors (models) defined on the Lawvere theory send the integer objects to the product objects of the models; e.g. for a functor R, the image $R(6)$ would be sent to a product of the following form.

$$R(\bullet) \times R(\bullet) \times R(\bullet) \times R(\bullet) \times R(\bullet) \times R(\bullet)$$

In the case of DNA, the idea is to copy the previous picture, but by adding enough information to be able to do genetics. If one looks at the type of pictures drawn by biologists to explain homologous recombination, alignment methods or even gene linkage, one can often see pictures of chromosomal patches subdivided in terms of selected and masked regions, as

[1] A small category equipped with a subset of its wide spans.

shown below.



These colored regions are obviously reminiscent of the term *chromo-some*[2] itself. The regional separations are also reminiscent of some sort of topology – or metric. If one tries to merge these topological and colored components with the type of atomic representation given in (2.1), we are likely to end up with the following type of picture.

$$(\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) \tag{2.2}$$

In the previous picture, the black nodes could indicate the colored regions while the white nodes could be parts of the chromosome that one wants to ignore (mask). For convenience, in the examples and illustrations of this article, we will only consider two colors (black & white), but the theory will work for more than two colors. Our set of colors will be encoded by pre-ordered sets, whose semantics will allow us to *select* and *cut*.

2.1. **Pre-ordered sets.** In this paper, the most basic notions of ordered set are expected to be known by the reader (e.g. partially ordered sets; totally (or linearly) ordered sets; pre-ordered sets ; see [15, Page 11]). However, because pre-orders play an important role later on, it was felt appropriate to recall their definition here.

A *pre-ordered set* consists of a set Ω and a binary relation $\leq$ on Ω satisfying the following logical implications.

1) (reflexivity) for every $x \in \Omega$, the relation $x \leq x$ holds;
3) (transitivity) for every $x, y, z \in \Omega$, if $x \leq y$ and $y \leq z$ hold, then so does $x \leq z$.

Example 2.1. The set $\{0, 1\}$ is a pre-ordered set if one sets $0 \leq 1$; $0 \leq 0$ and $1 \leq 1$. The resulting pre-ordered set is usually known as the *Boolean* pre-ordered set and the values 0 and 1 are usually denoted as `false` and `true`, respectively.

Remark 2.2 (Representation). Pre-ordered sets may happen to be sets of labels (or even sets of structures) instead of being sets of integers. In the case of the Boolean pre-ordered set given in Example 2.1, the labels `false` and `true` will sometimes be used instead of the integers 0 and 1, mainly for the sake of clarity.

Example 2.3. The set $\{0, 1\}$ could also be equipped with the discrete pre-order made of the reflexive relations $0 \leq 0$ and $1 \leq 1$ only.

Example 2.4. For every integer n, the n-fold Cartesian product $\{0, 1\}^{\times n}$ of the pre-ordered set given in Example 2.1 is equipped with a pre-order relation $\leq$ that compares two tuples in $\{0, 1\}^{\times n}$, say of the form $(x_1, \ldots, x_n) \leq (y_1, \ldots, y_n)$, if, and only if, the relation $x_i \leq y_i$ holds for every $1 \leq i \leq n$.

Example 2.5. The interval $[0, 1]$ is a pre-ordered set for the usual pre-order "being less than or equal to" defined on the set $\mathbb{R}$ of real numbers.

[2]meaning *color-body*

Remark 2.6 (Pre-order categories). A pre-ordered set is equivalently a category in which there exists at most one arrow between every pair of objects. In the sequel, a pre-ordered set will sometimes be called a *pre-order category* to emphasize its categorical nature.

2.2. **Finite sets of integers.** For every positive integer n, we will denote by $[n]$ the finite set of integers $\{1, 2, \ldots, n\}$. We will also let $[0]$ denote the empty set. In the sequel, for every non-negative integer n, the set $[n]$ will implicitly be equipped with the order associated with the set of integers (note that the restriction of this order on $[0]$ is the empty order).

2.3. **Segments.** Let $(\Omega, \preceq)$ denote a pre-ordered set. A *segment* over Ω consists of a pair of non-negative integers (n_1, n_0), an order-preserving[3] surjection $t : [n_1] \to [n_0]$ and a function $c : [n_0] \to \Omega$.

Remark 2.7 (Representation). Segments have all the necessary data to encode the type of pictures given in (2.2). For a segment (t, c) as defined above, the integer $[n_1]$ represents the range of elements composing the segment

$$n_1 = \ (\bullet\bullet\cdots\bullet)$$

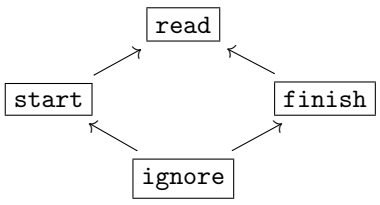
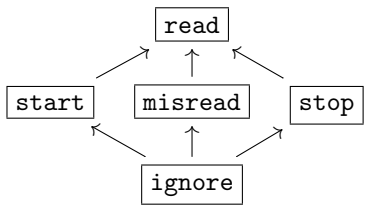
while the fibers $t^{-1}(1), \ldots, t^{-1}(n_0)$ of the surjection $t : [n_1] \to [n_0]$ gather these elements into patches (see the brackets below).

$$t = \ (\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet)(\bullet\bullet\cdots\bullet)(\bullet\bullet)$$

Finally, the different colors associated with the patches of the segment are specified by the map $c : [n_0] \to \Omega$. For instance, if we take Ω to be the Boolean pre-ordered set $\{\texttt{false} \leq \texttt{true}\}$ of Example 2.1 and we choose to associate the white color with the `false` value and the black color with the `true` value, then an identity of the form $c(1) = \texttt{false}$ will be represented by coloring all the elements of $[n_1]$ living in the fiber $t^{-1}(1)$ in white.

$$(t, c) = \ (\circ\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\cdots\bullet)(\bullet\bullet)$$

Note that if Ω contains more elements, then we need to use more colors (which can also be represented by numbers). These colors could also mean all sorts of things, including actions such as `ignore`, `read`, `start reading`, `stop reading`, `misread` (or `mutate`). The pre-order on the colors would then specify priorities or semantic inclusions between the different tasks associated with the colors (see below).

2 colors	4 colors	5 colors
$\{0, 1\}$	$\{0, 1, 2, 3\}$	$\{0, 1, 2, 3, 4\}$
		

Remark 2.8 (Notations). Note that the specification of the data n_1 and n_0 is redundant with the data of the function t and c. Later on, a segment will often be denoted as a pair (t, c) and, every so often, as an arrow $(t, c) : [n_1] \multimap [n_0]$.

Convention 2.9 (Domains, topologies & types)**.** For every segment $(t, c) : [n_1] \multimap [n_0]$, the data $[n_1]$ will be called the *domain* of (t, c), the data t will be called the *topology* of (t, c) and the data (n_1, n_0) will be called the *type* of (t, c). The type of a segment will always be specified as an arrow of the form $[n_1] \multimap [n_0]$.

Definition 2.10 (Homologous segments)**.** Two segments (t, c) and (t', c') over Ω will be said to be *homologous* if their topologies are equal.

[3]for every relation $x \leq y$ in the domain, the relation $t(x) \leq t(y)$ holds in the codomain.

Definition 2.11 (Quasi-homologous segments)**.** Two segments (t, c) and (t', c') over Ω will be said to be *quasi-homologous* if their domains are equal.

2.4. **Morphisms of segments.** Let $(\Omega, \preceq)$ be a pre-ordered set and $(t, c) : [n_1] \multimap [n_0]$ and $(t', c') : [n'_1] \multimap [n'_0]$ be two segments over Ω. A morphism of segments from (t, c) to (t', c') consists of

1) an order-preserving injection $f_1 : [n_1] \to [n'_1]$;
2) an order-preserving function $f_0 : [n_0] \to [n'_0]$;

such that the inequality $c' \circ f_0(i) \preceq c(i)$ holds for every $i \in [n_0]$ and the following diagram commutes.

$$\begin{tikzcd} {[n_1]} \arrow[r, two heads, "t"] \arrow[d, hook, "f_1"'] & {[n_0]} \arrow[d, "f_0"] \\ {[n'_1]} \arrow[r, two heads, "t'"] & {[n'_0]} \end{tikzcd}$$

It is easy to check that the class of morphisms of segments over Ω is stable under component-wise compositions and admits identities on every segment. We will denote by $\mathbf{Seg}(\Omega)$ the category whose objects are segments over Ω and whose arrows are morphisms between these.

From now on, we will regard the notations f_1 and f_0 given above as a conventional notationm for morphisms in $\mathbf{Seg}(\Omega)$. Below, we give several examples of typical morphism in $\mathbf{Seg}(\Omega)$ where Ω is taken to be the Boolean pre-ordered set of Example 2.1.

Example 2.12 (Locality)**.** If both components f_1 and f_0 are identities, then the inequality $c' \circ f_0 \preceq c$ 'decreases' the colors of the segment as illustrated below, on the left.

$$\begin{tikzcd} (\bullet\bullet\bullet)(\bullet\bullet)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\bullet) \arrow[d] \\ (\circ\circ\circ)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) \end{tikzcd} \qquad \begin{tikzcd} (\ldots)(\circ\circ\circ)(\ldots) \arrow[d] \\ (\ldots)(\bullet\bullet\bullet)(\ldots) \end{tikzcd}$$

Interpretation: This type of morphism tells us that one is able to select/cut local patches from a segment. When one has more than 2 colors (black & white), one is also able to prioritize certain patches via the order relation of the pre-ordered set Ω. Note that turning white regions into white ones, as shown above, on the right, is forbidden.

Example 2.13 (Metric)**.** If only the component f_1 is an identity morphism, then the component f_0 can merge the regions defining the topology.

$$\begin{tikzcd} (\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) \arrow[d] \\ (\circ\circ\circ\circ\circ)(\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ\circ) \end{tikzcd}$$

Interpretation: This type of morphism can be used to say that the way one parses the patches of a segment influences the way one parses the segment (e.g. from codons to genes), but the way one parses a segment does not necessarily reflect the way the patches are parsed (e.g from gene to codons).

Example 2.14 (Flexibility)**.** If the component f_1 is not an identity morphism, then the range of the segment increases. Below, we suppose that the identity $c' \circ f_0 = c$ holds.

$$\begin{tikzcd} (\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) \arrow[d] \\ (\bullet\bullet\bullet\bullet)(\bullet)(\circ\circ\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) \end{tikzcd}$$

Interpretation: This type of morphism allows one to insert particular nucleobases or spaces in the parsing of a segment. For instance, spaces become necessary if one wants to recombine segments that are not necessarily (quasi-)homologous. A morphism inserting a space would

then correspond to a choice of 'sequence alignment' in bioinformatics (this will be illustrated in Example 3.14).

Remark 2.15 (Initial object). For every pre-ordered set Ω, the segment (over Ω) of type $[0] \multimap [0]$ that is given by the obvious order-preserving surjection $! : \emptyset \to \emptyset$ and the canonical function $! : \emptyset \to \Omega$ is an initial object in $\mathbf{Seg}(\Omega)$.

2.5. **Pre-orders on homologous segments.** Let $(\Omega, \preceq)$ be a pre-ordered set and let $t : [n_1] \to [n_0]$ be a topology. The subcategory of $\mathbf{Seg}(\Omega)$ whose objects are the homologous segments of topology t and whose arrows are the morphisms of segments for which the components f_0 and f_1 are identities will be denoted by $\mathbf{Seg}(\Omega : t)$ and referred to as the *category of homologous segments (over Ω) of topology t.*

Proposition 2.16. *For every topology $t : [n_1] \to [n_0]$, the category $\mathbf{Seg}(\Omega : t)$ is a pre-order category*

Proof. The pre-order relations $(t, c) \leq (t, c')$ associated with $\mathbf{Seg}(\Omega : t)$ are induced by the following pre-order relations in $(\Omega, \preceq)$, which directly come from the definition of the arrows of $\mathbf{Seg}(\Omega : t)$.

$$c'(i) \preceq c(i)\,, \quad \forall i \in [n_0]$$

It is straightforward to see that this defines a reflexive and transitive relation. □

2.6. **Pre-orders on quasi-homologous segments.** Let $(\Omega, \preceq)$ be a pre-ordered set and let n_1 be a non-negative integer. The subcategory of $\mathbf{Seg}(\Omega)$ whose objects are the quasi-homologous segments of domain $[n_1]$ and whose arrows are the morphisms segments for which the component f_1 is an identity will be denoted by $\mathbf{Seg}(\Omega \,|\, n_1)$ and called the *category of quasi-homologous segments (over Ω) of domain n_1.*

Proposition 2.17. *If there exists a morphism $(t, c) \to (t', c')$ in $\mathbf{Seg}(\Omega \,|\, n_1)$, then it is the unique morphism of type $(t, c) \to (t', c')$ in $\mathbf{Seg}(\Omega)$*

Proof. Let $(\mathrm{id}, f_0) : (t, c) \to (t', c')$ be the morphism of the statement in $\mathbf{Seg}(\Omega \,|\, n_1)$ and let $(g_1, g_0) : (t, c) \to (t', c')$ be another morphism in $\mathbf{Seg}(\Omega)$. Because g_1 is an order-preserving inclusion of type $[n_1] \to [n_1]$, it must be an identity, so that the identity $g_0 \circ t = t'$ holds. On the other hand, the identity $f_0 \circ t = t'$ also holds, which means that $g_0 \circ t = f_0 \circ t$. Because t is an epimorphism, the identity $g_0 = f_0$ must hold. □

Remark 2.18 (Preorder category). Proposition 2.17 implies that the category $\mathbf{Seg}(\Omega \,|\, n_1)$ of quasi-homologous segments is a pre-order category.

Remark 2.19 (Zero domain). The category $\mathbf{Seg}(\Omega \,|\, 0)$ of quasi-homologous segments of empty domain is a terminal category whose only object is the initial object of $\mathbf{Seg}(\Omega)$.

2.7. **Cones.** Recall that a *cone* in a category $\mathcal{C}$ consists of an object X in $\mathcal{C}$, a small category $\mathtt{A}$, a functor $U : \mathtt{A} \to \mathcal{C}$ and a natural transfomation $\Delta_{\mathtt{A}}(X) \Rightarrow U$ where $\Delta_{\mathtt{A}}(X)$ denotes the constant functor $\mathtt{A} \to \mathbf{1} \to \mathcal{C}$ mapping every object in $\mathtt{A}$ to the object X in $\mathcal{C}$.

2.8. **Chromologies.** A *chromology* is a pre-ordered set $(\Omega, \preceq)$ that is equipped, for every non-negative integer n, with a set $D[n]$ of cones in the category $\mathbf{Seg}(\Omega \,|\, n)$. A chromology as above will later be denoted as a pair (Ω, D).

Remark 2.20 (Future examples). In section 3.3, we will see several examples of chromologies, which will be used throughout this article.

2.9. **Logical systems.** We will speak of a *logical system* to refer to a category $\mathcal{C}$ that is equipped with a subclass of its cones $\mathcal{W}$ (see section 2.7).

Remark 2.21 (Limit sketch). The difference between a logical system and a limit sketch is that the latter is defined as a small category that is equipped with a subset of its cones. A logical system is also meant to be the codomain of a functor whose domain is a limit sketch.

2.10. **Pedigrads.** Pedigrads are algebraic structures that model the logical rules of chromologies. Their name refers to the concept of 'pedigree' used in genetics. Let (Ω, D) be a chromology and $(\mathcal{C}, \mathcal{W})$ be a logical system. A *pedigrad* in $(\mathcal{C}, \mathcal{W})$ is a functor $\mathbf{Seg}(\Omega) \to \mathcal{C}$ sending, for every non-negative integer n, the cones in $D[n]$ to cones in $\mathcal{W}$.

Convention 2.22 ($\mathcal{W}$-pedigrads)**.** Because we will often consider the same category $\mathcal{C}$ for different classes of cones $\mathcal{W}$, we will sometimes refer to a pedigrad in $(\mathcal{C}, \mathcal{W})$ as a $\mathcal{W}$-pedigrad.

2.11. **Morphisms of pedigrads.** Recall that, for every pair of categories $\mathcal{C}$ and $\mathcal{D}$, the notation $[\mathcal{C}, \mathcal{D}]$ denotes the category whose objects are functors $\mathcal{C} \to \mathcal{D}$ and whose arrows are natural transformations in $\mathcal{D}$ over $\mathcal{C}$. Let (Ω, D) be a chromology and $(\mathcal{C}, \mathcal{W})$ be a logical system. A *morphism of pedigrads* from a pedigrad $A : \mathbf{Seg}(\Omega) \to \mathcal{C}$ in $(\mathcal{C}, \mathcal{W})$ for (Ω, D) to a pedigrad $B : \mathbf{Seg}(\Omega) \to \mathcal{C}$ in $(\mathcal{C}, \mathcal{W})$ for (Ω, D) is an arrow $A \Rightarrow B$ in the category $[\mathbf{Seg}(\Omega), \mathcal{C}]$.

Convention 2.23 (Category of pedigrads)**.** The full subcategory of $[\mathbf{Seg}(\Omega), \mathcal{C}]$ whose objects are the pedigrads in a logical system $(\mathcal{C}, \mathcal{W})$ for a chromology (Ω, D) will be called *the category of pedigrads in* $(\mathcal{C}, \mathcal{W})$ *for* (Ω, D).

3. Examples of pedigrads in sets

The goal of this section is to construct two canonical classes of pedigrads that take their values in the category **Set** of sets and functions (see Propositions 3.23 & 3.28). Throughout the section, we shall also let (E, ε) be a fixed pointed set and $(\Omega, \preceq)$ be a pre-ordered set.

3.1. **Truncation functors.** In this section, we define a truncation operation, which will turn out to be very useful for constructing pedigrads.

Definition 3.1 (Truncation)**.** For every segment $(t, c) : [n_1] \multimap [n_0]$ over Ω and element $b \in \Omega$, we will denote by $\mathrm{Tr}_b(t, c)$ the subset $\{i \in [n_1] \mid b \preceq c \circ t(i)\}$ of $[n_1]$. This is the set of all elements in $[n_1]$ whose images via $c \circ t$ is greater than or equal to b in Ω.

Example 3.2 (Truncation)**.** Let $(\Omega, \preceq)$ be the Boolean pre-ordered set $\{0 \leq 1\}$. If we consider a segment over Ω, as given below, on the left, the operation Tr_b for which b is taken to be equal to 1 will select all the integers in the domain of (t, c) that are associated with black nodes while the operation Tr_b for which b is taken to be equal to 0 will select all the integers in the domain of (t, c).

$$(t, c) = (\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\circ\circ\circ\circ\circ)(\bullet\bullet\bullet)(\circ) \qquad \begin{array}{l} \mathrm{Tr}_1(t, c) = \{1, 2, 3, 6, 7, 8, 9, 15, 16, 17\} \\ \mathrm{Tr}_0(t, c) = [18] \end{array}$$

Definition 3.3 (Sub-objects)**.** For every non-negative integer n, we will speak of a *sub-object* of $[n]$ to refer to a subset of $[n]$. A *morphism of sub-objects of* $[n]$ is an inclusion of sets between the two sub-objects.

Example 3.4 (Truncation operations and sub-objects)**.** Let $(\Omega, \preceq)$ be the Boolean pre-ordered set $\{0 \leq 1\}$. If we consider the morphism of segments of Example 2.13, which is recalled below, on the left, we can see that the truncation operation Tr_1 gives, on the right, two sub-objects of the domain [18] that we can relate via a morphism of sub-objects.

$$\begin{array}{ccc} (t, c) = (\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) & & \mathrm{Tr}_1(t, c) = \{1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14\} \\ \downarrow & & \rotatebox{90}{$\subseteq$} \\ (t', c') = (\circ\circ\circ\circ\circ)(\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ\circ) & & \mathrm{Tr}_1(t', c') = \{6, 7, 8, 9, 10, 11, 12, 13, 14\} \end{array}$$

The fact that a morphism of segments of the form $(t, c) \to (t', c')$ gives rise to an inclusion $\mathrm{Tr}_1(t', c') \subseteq \mathrm{Tr}_1(t, c)$ is explained by Proposition 3.5.

Proposition 3.5. *Let* $(f_1, f_0) : (t, c) \to (t', c')$ *be a morphism in* $\mathbf{Seg}(\Omega)$*. If the relation* $f_1(i) \in \mathrm{Tr}_b(t', c')$ *holds, then so does the relation* $i \in \mathrm{Tr}_b(t, c)$*.*

Proof. Recall that, by definition of a morphism in $\mathbf{Seg}(\Omega)$, the inequality $c' \circ f_0 \preceq c$ holds. Now, if the relation $f_1(i) \in \mathrm{Tr}_b(t', c')$ holds, then so do the following pre-order relations.

$$b \preceq c' \circ t' \circ f_1(i) = c' \circ f_0 \circ t(i) \preceq c \circ t(i)$$

By transitivity, we obtain the inequality $b \preceq c \circ t(i)$, so that i must be in $\mathrm{Tr}_b(t, c)$. □

Proposition 3.6. *For every element $b \in \Omega$ and non-negative integer n_1, the mapping $(c, t) \mapsto \mathrm{Tr}_b(t, c)$ extends to a functor $\mathrm{Tr}_b : \mathbf{Seg}(\Omega \,|\, n_1) \to \mathbf{Set}^{\mathrm{op}}$, which factorizes through the opposite category of sub-objects of $[n_1]$.*

Proof. By definition, for every segment (c, t) in $\mathbf{Seg}(\Omega \,|\, n_1)$, the set $\mathrm{Tr}_b(t, c)$ is a subset of $[n_1]$. For every morphism $(\mathrm{id}, f_0) : (t, c) \to (t, c')$ in $\mathbf{Seg}(\Omega \,|\, n_1)$, Proposition 3.5 shows that there is an inclusion $\mathrm{Tr}_b(t', c') \subseteq \mathrm{Tr}_b(t, c)$. Since the opposite category of sub-objects of $[n_1]$ is a pre-order category, the statement follows. □

In fact, Proposition 3.6 hides a more general construction if one allows the consideration of the category $\mathbf{Set}_*$ of pointed sets and point-preserving maps (see Example 3.8). Recall that there is an adjunction

$$\begin{tikzcd} \mathbf{Set} \arrow[r, shift left, "\mathbb{F}"] \arrow[r, phantom, "\perp"] & \mathbf{Set}_* \arrow[l, shift left, "\mathbb{U}"] \end{tikzcd}$$

whose right adjoint $\mathbb{U} : \mathbf{Set}_* \to \mathbf{Set}$ forgets the pointed structure (i.e. $\mathbb{U} : (X, p) \mapsto X$) and whose left adjoint $\mathbb{F} : \mathbf{Set} \to \mathbf{Set}_*$ maps a set X to the obvious pointed set $(X + \{*\}, *)$ and maps a function $f : X \to Y$ to the coproduct map $f + \{*\} : X + \{*\} \to Y + \{*\}$.

Proposition 3.7. *For every element $b \in \Omega$, the mapping $(c, t) \mapsto \mathbb{F}\mathrm{Tr}_b(t', c')$ extends to a functor $\mathrm{Tr}_b^* : \mathbf{Seg}(\Omega) \to \mathbf{Set}_*^{\mathrm{op}}$ mapping every function $(f_1, f_0) : (t, c) \to (t', c')$ in $\mathbf{Seg}(\Omega)$ to the following map of pointed sets.*

$$\begin{array}{rcccl} \mathrm{Tr}_b^*(f_1, f_0) & : & \mathbb{F}\mathrm{Tr}_b(t', c') & \to & \mathbb{F}\mathrm{Tr}_b(t, c) \\ & & j & \mapsto & i \qquad \textit{if } \exists i \in \mathrm{Tr}_b(t, c) \,:\, j = f_1(i); \\ & & j & \mapsto & * \qquad \textit{otherwise.} \end{array}$$

Proof. Follows from Proposition 3.5. □

Example 3.8 (Truncation operations and pointed sets)**.** Let $(\Omega, \preceq)$ be the Boolean pre-ordered set $\{0 \leq 1\}$. If we consider the morphism of segments of Example 2.14, which is further specifed below, on the left, by using adequate labeling to show how the first segment is mapped to the second one, we can see that the truncation operation Tr_1, displayed on the right, forces us to consider a map of pointed sets.

$$\begin{array}{ccccc} (\overset{123}{\bullet\bullet\bullet})(\overset{45}{\circ\circ})(\overset{6789}{\bullet\bullet\bullet\bullet})(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) & & \{1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14\} & & ? \\ \downarrow & & \uparrow & & \uparrow \\ (\overset{123*}{\bullet\bullet\bullet\bullet})(\overset{*}{\bullet})(\overset{45**}{\circ\circ\circ\circ})(\overset{6789}{\bullet\bullet\bullet\bullet})(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) & & \{1, 2, 3, 10, 11, 12, \ldots, 16, 17, 18\} & \cup & \{4, 5\} \end{array}$$

Proposition 3.9. *For every element $b \in \Omega$ and non-negative integer n_1, the following diagram commutes.*

$$\begin{tikzcd} \mathbf{Seg}(\Omega \,|\, n_1) \arrow[r, "\subseteq"] \arrow[d, "\mathrm{Tr}_b"'] & \mathbf{Seg}(\Omega) \arrow[d, "\mathrm{Tr}_b^*"] \\ \mathbf{Set}^{\mathrm{op}} \arrow[r, "\mathbb{F}^{\mathrm{op}}"'] & \mathbf{Set}_*^{\mathrm{op}} \end{tikzcd}$$

Proof. By definition, if we restrict the functor $\mathrm{Tr}_b^* : \mathbf{Seg}(\Omega) \to \mathbf{Set}_*^{\mathrm{op}}$ to the subcategory $\mathbf{Seg}(\Omega \,|\, n_1) \hookrightarrow \mathbf{Seg}(\Omega)$, then every morphism $(t, c) \leq (t, c')$ in $\mathbf{Seg}(\Omega \,|\, n_1)$ is sent to the

following map in $\mathbf{Set}_*$.

$$\begin{array}{rcccl} \mathrm{Tr}_b^*(f_1, f_0) & : & \mathbb{F}\mathrm{Tr}_b(t, c') & \to & \mathbb{F}\mathrm{Tr}_b(t, c) & \\ & & j & \mapsto & j & j \in \mathrm{Tr}_b(t, c') \\ & & * & \mapsto & * & \text{otherwise.} \end{array}$$

This means that the restriction of Tr_b^* on $\mathbf{Seg}(\Omega \,|\, n_1)$ can be retrieved from the application of the functor $\mathbb{F}$ on the images of Tr_b. □

3.2. **Example of pedigrads in sets.** In this section, we construct a collection of functors $\mathbf{Seg}(\Omega) \to \mathbf{Set}$ by using the pointed set (E, ε) and a parameter in $b \in \Omega$ (see Definition 3.11). Later on, we will define various classes of cones $\mathcal{W}$ in $\mathbf{Set}$ for which these functors are $\mathcal{W}$-pedigrad.

Convention 3.10 (Notation)**.** In the sequel, the hom-set of a category $\mathcal{C}$ from an object X to an object Y will be denoted as $\mathcal{C}(X, Y)$. For instance, the set of functions from a set X to a set Y will be denoted by $\mathbf{Set}(X, Y)$. Recall that, for any category $\mathcal{C}$, the hom-sets give rise to a functor $\mathcal{C}(_, _) : \mathcal{C}^{\mathrm{op}} \times \mathcal{C} \to \mathbf{Set}$ called the *hom-functor* [15, page 27].

Definition 3.11 (Canonical pedigrads)**.** For every element $b \in \Omega$, we will denote by E_b^ε the functor $\mathbf{Seg}(\Omega) \to \mathbf{Set}$ defined as the composition of the following functors.

$$\mathbf{Seg}(\Omega) \xrightarrow{\mathrm{Tr}_b^*} \mathbf{Set}_*^{\mathrm{op}} \xrightarrow{\mathbf{Set}_*(_,(E,\varepsilon))} \mathbf{Set}$$

Remark 3.12. For every object (t, c) in $\mathbf{Seg}(\Omega)$, an element in $E_b^\varepsilon(t, c)$ can be seen as a function of the form $\mathrm{Tr}_b(t, c) \to E$ according to the following series of bijections.

$$\begin{aligned} E_b^\varepsilon(t, c) &= \mathbf{Set}_*(\mathrm{Tr}_b^*(t, c), (E, \varepsilon)) \\ &= \mathbf{Set}_*(\mathbb{F}\mathrm{Tr}_b(t, c), (E, \varepsilon)) && \text{(Def. of } \mathrm{Tr}_b^*) \\ &\cong \mathbf{Set}(\mathrm{Tr}_b(t, c), \mathbb{U}(E, \varepsilon)) && (\mathbb{F} \dashv \mathbb{U}) \\ &= \mathbf{Set}(\mathrm{Tr}_b(t, c), E) && \text{(Def. of } \mathbb{U}) \end{aligned}$$

Because the set $\mathrm{Tr}_b(t, c)$ is equipped with the natural order of natural numbers, we will represent an element in $E_b^\varepsilon(t, c)$ as a word of elements in E (see Example 3.13).

Example 3.13 (Objects)**.** Suppose that Ω denotes the Boolean pre-ordered set $\{0 \leq 1\}$ and let (E, ε) be the pointed set $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}, \varepsilon\}$. If we consider the segment

$$(c, t) = \ (\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ)$$

then the set $E_1^\varepsilon(t, c)$ (where $b = 1$) will contain the following words (which have been parenthesized for clarity), among many others.

$$\begin{gathered} (\mathtt{AG}\varepsilon)(\mathtt{TCAA})(\mathtt{TAGG}\varepsilon) \\ (\mathtt{GT}\varepsilon)(\varepsilon\varepsilon\varepsilon\mathtt{C})(\mathtt{AGTAC}) \\ (\mathtt{TAA})(\mathtt{GATC})(\mathtt{AGTTT}) \\ \text{etc.} \end{gathered}$$

Example 3.14 (Morphisms)**.** Suppose that Ω denotes the Boolean pre-ordered set $\{0 \leq 1\}$ and let (E, ε) be the pointed set $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}, \varepsilon\}$. If we consider the morphism of segments given below, in which we use adequate labeling to show how the first segment is included in the second one,

$$(\overset{1}{\bullet}\overset{2}{\bullet}\overset{3}{\bullet})(\overset{4}{\circ}\overset{5}{\circ})(\overset{6}{\bullet}\overset{7}{\bullet}\overset{8}{\bullet}\overset{9}{\bullet})(\overset{10}{\bullet}\ \overset{11}{\bullet}) \ \to \ (\overset{1}{\bullet}\overset{2}{\bullet}\overset{3}{\bullet}\overset{*}{\bullet}\overset{*}{\bullet})(\overset{4}{\circ}\overset{5}{\circ}\overset{*}{\circ})(\overset{6}{\bullet}\overset{7}{\bullet}\overset{8}{\bullet}\overset{9}{\bullet})(\overset{*}{\bullet})(\overset{10}{\circ}\ \overset{11}{\circ})$$

then the function obtained from the application of E_1^ε will have the following mapping rules.

$$\begin{array}{ccc} (\mathtt{AG}\varepsilon)(\mathtt{TCAA})(\mathtt{GC}) & \mapsto & (\mathtt{AG}\varepsilon\varepsilon\varepsilon)(\mathtt{TCAA})(\varepsilon) \\ (\mathtt{GT}\varepsilon)(\varepsilon\varepsilon\varepsilon\mathtt{C})(\mathtt{TA}) & \mapsto & (\mathtt{GT}\varepsilon\varepsilon\varepsilon)(\varepsilon\varepsilon\varepsilon\mathtt{C})(\varepsilon) \\ (\mathtt{TAA})(\mathtt{GATC})(\mathtt{AA}) & \mapsto & (\mathtt{TAA}\varepsilon\varepsilon)(\mathtt{GATC})(\varepsilon) \\ & \text{etc.} & \end{array}$$

If one restricts oneself to morphisms that only insert symbols ε and do not turn any black node into white ones, then the mappings associated with this type of morphism can be assimilated to the gap insertion operations used in sequence alignment algorithms (see [25] or [20, Chap. 1, sec. 5]) in order to compare two sequences of different lengths [30, 21].

$$(\overset{1}{\bullet}\overset{2}{\bullet}\overset{3}{\bullet})(\overset{4}{\circ}\overset{5}{\circ})(\overset{6}{\bullet}\overset{7}{\bullet}\overset{8}{\bullet}\overset{9}{\bullet})(\overset{10}{\bullet}\ \overset{11}{\bullet}) \ \to\ (\overset{1}{\bullet}\overset{2}{\bullet}\overset{3}{\bullet})(\overset{4}{\circ}\overset{5}{\circ}\overset{*}{\circ})(\overset{6}{\bullet}\overset{7}{\bullet}\overset{*}{\bullet}\overset{*}{\bullet}\overset{8}{\bullet}\overset{9}{\bullet})(\overset{*}{\bullet}\overset{10}{\bullet}\ \overset{11}{\bullet})$$

$$\begin{array}{ccc} (\mathtt{GAC})(\mathtt{ATTC})(\mathtt{CT}) & \mapsto & (\mathtt{GAC})(\mathtt{AT}\varepsilon\varepsilon\mathtt{TC})(\varepsilon\mathtt{CT}) \\ & \text{etc.} & \end{array}$$

Proposition 3.15. *For every domain $[n_1]$, the restriction of the functor $E_b^\varepsilon : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ on $\mathbf{Seg}(\Omega \,|\, n_1)$ is isomorphic to the functor $\mathbf{Set}(\mathrm{Tr}_b(_), E) : \mathbf{Seg}(\Omega \,|\, n_1) \to \mathbf{Set}$. In other words, the following diagram commutes up to an isomorphism of functors.*

$$\begin{tikzcd} \mathbf{Seg}(\Omega \,|\, n_1) \arrow[r, "\subseteq"] \arrow[d, "\mathbf{Tr}_b"'] & \mathbf{Seg}(\Omega) \arrow[d, "E_b^\varepsilon"] \\ \mathbf{Set}^{\mathrm{op}} \arrow[r, "\mathbf{Set}(_,E)"'] & \mathbf{Set} \end{tikzcd}$$

Proof. Note that the following series of isomorphisms hold on $\mathbf{Seg}(\Omega \,|\, n_1)$.

$$\begin{aligned} E_b^\varepsilon(_) &= \mathbf{Set}_*\Big(\mathrm{Tr}_b^*(_), (E, \varepsilon)\Big) \\ &= \mathbf{Set}_*\Big(\mathbb{F}\mathrm{Tr}_b(_), (E, \varepsilon)\Big) && \text{(Proposition 3.9)} \\ &\cong \mathbf{Set}\Big(\mathrm{Tr}_b(_), \mathbb{U}(E, \varepsilon)\Big) && (\mathbb{F} \dashv \mathbb{U}) \\ &= \mathbf{Set}\Big(\mathrm{Tr}_b(_), E\Big) && \text{(Def. of } \mathbb{U}) \end{aligned}$$

Because these isomorphisms are natural on $\mathbf{Seg}(\Omega \,|\, n_1)$, the statement follows. □

3.3. **Distributive and exactly distributive chromologies.** Let $(\Omega, \preceq)$ be a pre-ordered set, b be an element in Ω, A be a small category and $\rho : \Delta_A(\tau) \Rightarrow \theta$ be a cone in $\mathbf{Seg}(\Omega \,|\, n)$ for some non-negative integer n. Note that the application of the truncation functor $\mathrm{Tr}_b : \mathbf{Seg}(\Omega \,|\, n) \to \mathbf{Set}^{\mathrm{op}}$ on the cone ρ gives rise to a cocone in $\mathbf{Set}$ as follows.

$$\mathrm{Tr}_b(\rho) : \mathrm{Tr}_b\theta \Rightarrow \Delta_A \circ \mathrm{Tr}_b(\tau) \tag{3.1}$$

According to Proposition 3.6, this cocone can be seen as a diagram in the category of subobjects of $[n]$. It follows that the colimit adjoint of (3.1) in $\mathbf{Set}$ can be factorized, via an epi-mono factorization, through the union of sets $\bigcup_{a \in A} \mathrm{Tr}_b\theta(a)$, as shown below.

$$\begin{tikzcd} \mathrm{colim}_A \mathrm{Tr}_b\theta \arrow[r, "e", "\text{epi.}"'] \arrow[rr, bend left, "\lim_A \mathrm{Tr}_b(\rho)"] & \bigcup_{a \in A} \mathrm{Tr}_b\theta(a) \arrow[r, "m", "\text{mono.}"'] & \mathrm{Tr}_b(\tau) \end{tikzcd} \tag{3.2}$$

Definition 3.16 (Distributive cones)**.** A cone of the form $\rho : \Delta_A(\tau) \Rightarrow \theta$ in $\mathbf{Seg}(\Omega \,|\, n)$ will be said to be *b-distributive* if the arrow m of (3.2) is an epimorphism (hence an identity).

Definition 3.17 (Exactly distributive cones)**.** A cone of the form $\rho : \Delta_A(\tau) \Rightarrow \theta$ in $\mathbf{Seg}(\Omega \,|\, n)$ will be said to be *exactly b-distributive* if it is b-distributive and the arrow e of (3.2) is a monomorphism (hence an bijection).

Example 3.18 (Homologous cones)**.** Let Ω denote the Boolean pre-ordered set $\{0 \leq 1\}$. We now give examples of 1-distributive and exactly 1-distributive cones taken in one of the subcategories of homologous segments of $\mathbf{Seg}(\Omega)$. We quickly describe the form of 0-distributive cones at the end of this example.

First, we can give the following three inequalities as an example of 1-distributive cone in one of the pre-order categories $\mathbf{Seg}(\Omega : t)$ for the obvious topology t of domain [18].

$$\begin{aligned}
(\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) &\leq (\circ\circ\circ)(\circ\circ)(\bullet\bullet\bullet\bullet)(\circ\circ\circ\circ\circ)(\circ\circ\circ)(\circ)\\
(\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) &\leq (\circ\circ\circ)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ)\\
(\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) &\leq (\bullet\bullet\bullet)(\circ\circ)(\circ\circ\circ\circ)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ)
\end{aligned}$$

From a biological point of view, this type of cone could be used to specify various ways of selecting a set of codons or genes (depending on the scale).

For their part, exactly 1-distributive cones cannot have common black patches that are not related via their diagram A. Below is an example of such a cone whose diagram is discrete.

$$\begin{aligned}
(\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) &\leq (\circ\circ\circ)(\circ\circ)(\bullet\bullet\bullet\bullet)(\circ\circ\circ\circ\circ)(\circ\circ\circ)(\circ)\\
(\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) &\leq (\circ\circ\circ)(\circ\circ)(\circ\circ\circ\circ)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ)\\
(\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) &\leq (\bullet\bullet\bullet)(\circ\circ)(\circ\circ\circ\circ)(\circ\circ\circ\circ\circ)(\circ\circ\circ)(\circ)
\end{aligned}$$

From a biological point of view, this type of cone could be used to specify which codon or gene (depending on the scale) is separated from the others during homologous recombination.

On the other hand, the class of 0-distributive cones in $\mathbf{Seg}(\Omega : t)$ contains all the cones of $\mathbf{Seg}(\Omega : t)$ while the class of exactly 0-distributive cones in $\mathbf{Seg}(\Omega : t)$ is equal to the set of identities (seen as one-arrow cones) in $\mathbf{Seg}(\Omega : t)$.

Example 3.19 (Quasi-homologous cones)**.** Let Ω denote the Boolean pre-ordered set $\{0 \leq 1\}$. In this example, we give 1-distributive and exactly 1-distributive cones in one of the categories of quasi-homologous segments of $\mathbf{Seg}(\Omega)$. For instance, the three (obvious) arrows given below (in which the bracketed elements $(\bullet)$ and $(\circ)$ have been shorten, on the left side, to the symbols $\bullet$ and $\circ$, respectively) form a single 1-distributive cone in $\mathbf{Seg}(\Omega \,|\, 18)$.

$$\begin{aligned}
\bullet\bullet\bullet\circ\circ\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\circ\circ\circ\circ &\leq (\circ\circ)(\circ)(\circ\circ)(\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ\circ)(\circ\circ\circ)(\circ)\\
\bullet\bullet\bullet\circ\circ\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\circ\circ\circ\circ &\leq (\circ\circ\circ)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ)\\
\bullet\bullet\bullet\circ\circ\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\circ\circ\circ\circ &\leq (\bullet\bullet\bullet)(\circ\circ)(\circ\circ\circ\circ)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ)
\end{aligned}$$

For their part, exactly 1-distributive cones cannot have common black patches that are not related via their diagram A. Below is an example of such a cone whose diagram is discrete.

$$\begin{aligned}
\bullet\bullet\bullet\circ\circ\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\circ\circ\circ\circ &\leq (\circ\circ)(\circ)(\circ\circ)(\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ\circ)(\circ\circ\circ)(\circ)\\
\bullet\bullet\bullet\circ\circ\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\circ\circ\circ\circ &\leq (\circ\circ\circ)(\circ\circ)(\circ\circ\circ\circ)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ)\\
\bullet\bullet\bullet\circ\circ\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\circ\circ\circ\circ &\leq (\bullet\bullet\bullet)(\circ\circ)(\circ\circ)(\circ\circ)(\circ\circ\circ\circ\circ)(\circ\circ\circ)(\circ)
\end{aligned}$$

The difference between the cones given in Example 3.18 and those given above is that the ones given above specify operations that act on sequences that are, *a priori*, not equipped with any particular topology. See the difference in this following example. Suppose that E_0 denotes an enzyme that is to cut a DNA sequence codon by codon. Either one feeds E_0 with a sequence whose topology is already specified, say $(\mathtt{ATC})(\mathtt{GA})$, and E_0 returns $\mathtt{ATC}$ and $\mathtt{GA}$ or one feeds E_0 with a sequence of the form $\mathtt{ATCGA}$ and only the specification of a particular cone would allow us to know what E_0 returns.

Definition 3.20 (Distributive chromologies)**.** For every element $b \in \Omega$, a chromology (Ω, D) will be said to be *b-distributive* if all its cones in D are b-distributive.

Definition 3.21 (Exactly distributive chromologies)**.** For every element $b \in \Omega$, a chromology (Ω, D) will be said to be *exactly b-distributive* if all its cones in D are exactly b-distributive.

3.4. **Logical systems for pedigrads in sets.** In this section, we show that the functor defined in Definition 3.11 can be seen as a pedigrad for a certain logical system in **Set**.

Definition 3.22 (Logical systems of monomorphisms)**.** We will denote by $\mathcal{W}^{\text{inj}}$ the set of cones $\Delta_A(X) \Rightarrow F$ in **Set** whose limit adjoints $X \to \lim_A F$ are injections.

Proposition 3.23. *For every element b in Ω and b-distributive chromology (Ω, D), the functor $E_b^\varepsilon : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ is a $\mathcal{W}^{inj}$-pedigrad for (Ω, D).*

Proof. Let $\rho : \Delta_A(\tau) \Rightarrow \theta$ be a cone in $D[n_1]$ for some given non-negative integer n_1. By assumption on (Ω, D) and Definition 3.16, the canonical arrow

$$\text{colim}_A \text{Tr}_b \theta \to \text{Tr}_b(\tau)$$

must be an epimorphism, so that its image via the functor $\mathbf{Set}(_, E) : \mathbf{Set}^{\text{op}} \to \mathbf{Set}$ is an injection. By Proposition 3.15 and the usual definition of colimits in **Set**, the resulting injection is (naturally) isomorphic to the following canonical arrow.

$$E_b^\varepsilon(\tau) \to \lim_A E_b^\varepsilon \circ \theta$$

This precisely shows that $E_b^\varepsilon : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ is a $\mathcal{W}^{\text{inj}}$-pedigrad for (Ω, D). □

Definition 3.24 (Sub-functors)**.** Recall that a *sub-functor* of a functor $G : D \to \mathbf{Set}$ is a functor $F : D \to \mathbf{Set}$ such that

1) for every object d in D, the set $F(d)$ is a subset in $G(d)$;
2) for every morphism $f : d \to d'$ in D, the function $F(f) : F(d) \to F(d')$ makes the following diagram commute.

$$\begin{tikzcd} F(d) \arrow[r, "\subseteq"] \arrow[d, "F(f)"'] & G(d) \arrow[d, "G(f)"] \\ F(d') \arrow[r, "\subseteq"'] & G(d') \end{tikzcd}$$

In this case, we will write $F \subseteq G$ to mean that F is a sub-functor of G.

Remark 3.25 (Sub-functors are pedigrads). Every sub-functor $F \subseteq E_b^\varepsilon$ is a $\mathcal{W}^{\text{inj}}$-pedigrad for any b-distributive chromology (Ω, D). This follows from the commutativity of the following diagram for every cone $\rho : \Delta_A(\tau) \Rightarrow \theta$ in D, which forces the left-most vertical arrow to be an injection.

$$\begin{tikzcd} F(\tau) \arrow[r, "\subseteq"] \arrow[d] & E_b^\varepsilon(\tau) \arrow[d, "\text{inj.}"] \\ \lim_A F \circ \theta \arrow[r] & \lim_A E_b^\varepsilon \circ \theta \end{tikzcd}$$

Example 3.26 (Living beings)**.** Let $(\Omega, \preceq)$ be the pre-ordered set $\{0 \leq 1\}$ and (E, ε) be the pointed set $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}, \varepsilon\}$. We will suppose that Ω is equipped with a 1-distributive chromology structure (Ω, D). The sub-functors of $E_1^\varepsilon : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ could be seen as structures containing the genomes of living beings whose genetic codes are encoded with DNA. Precisely, the fact that the cones of D are sent to $\mathcal{W}^{\text{inj}}$ means that the DNA strands

are *uniquely determined* by the patches living in the codomains of the arrows defining the cones of D.

$$\begin{tikzcd}[column sep=small] & \texttt{GGATACCGATTA} \arrow[dl, bend right] \arrow[d] \arrow[dr, bend left] & \\ (___)(\texttt{TAC})(_____) & (\texttt{GGA})(___)(_____) & (___)(___)(\texttt{CGATTA}) \end{tikzcd}$$

Note that if two patches S_1 and S_2 live in a subfunctor $F \subseteq E_1^\varepsilon$, then their concatenation $S_1 \cdot S_2$ might not exist in the sub-functor F (this makes sense with the fact that if two genes are present in an individual X, then the cutting and re-gluing of their internal patches might not exist in X). However, the ability of concatenating segments can turn out to be useful if one wants to deal with haplogroups and homologous recombination. To be able to concatenate DNA strands, we will need the type of pedigrad induced by the logical systems defined in Definition 3.27.

Definition 3.27 (Logical systems of bijections)**.** We will denote by $\mathcal{W}^{\text{bij}}$ the set of cones $\Delta_A(X) \Rightarrow F$ in **Set** whose limit adjoints $X \to \lim_A F$ are bijections.

Proposition 3.28. *For every element b in Ω and exactly b-distributive chromology (Ω, D), the functor $E_b^\varepsilon : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ is a $\mathcal{W}^{bij}$-pedigrad for (Ω, D).*

Proof. Let $\rho : \Delta_A(\tau) \Rightarrow \theta$ be a cone in $D[n_1]$ for some given non-negative integer n_1. By assumption on (Ω, D) and Definitions 3.16 & 3.17, the canonical arrow

$$\text{colim}_A \text{Tr}_b \theta \to \text{Tr}_b(\tau)$$

must be both an epimorphism and a monomorphism (and, in fact, an isomorphism), so that its image via the functor $\mathbf{Set}(_, E) : \mathbf{Set}^{\text{op}} \to \mathbf{Set}$ is a bijection. By Proposition 3.15 and the usual definition of colimits in **Set**, the resulting bijection is (naturally) isomorphic to the following canonical arrow.

$$E_b^\varepsilon(\tau) \to \lim_A E_b^\varepsilon \circ \theta$$

This precisely shows that $E_b^\varepsilon : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ is a $\mathcal{W}^{\text{bij}}$-pedigrad for (Ω, D). □

Example 3.29 (Algebraic ring of DNA)**.** The idea behind $\mathcal{W}^{\text{bij}}$-pedigrads is that they allow one to model all those operations that an engineer might want to do (such as homologous recombination, CRISPR [23], DNA sequencing, alignment methods, etc.), but that could not exist in a living being. Therefore $\mathcal{W}^{\text{bij}}$-pedigrads should be seen as universes in which it is possible to think about DNA, instead of being seen as models containing the genome of a particular set of living beings.

Example 3.30 (Duplication)**.** Let $(\Omega, \preceq)$ be the pre-ordered set $\{0 \leq 1\}$ and (E, ε) be the pointed set $\{\texttt{A}, \texttt{C}, \texttt{G}, \texttt{T}, \varepsilon\}$. In this example, we illustrate how flexible $\mathcal{W}^{\text{bij}}$-pedigrads are by constructing what one could see as a duplication mutation, which is a type of mutation that is responsible for triggering certain cancers [24]. To do so, first note that the category $\mathbf{Seg}(\Omega)$ contains two morphisms as follows (in which the labeled elements $\bullet$ specify how the domain is sent to the codomain).

$$(\overset{1}{\bullet}\overset{2}{\bullet}\overset{3}{\bullet}) \xrightarrow{f_1} (\overset{1}{\bullet}\overset{2}{\bullet}\overset{3}{\bullet})(\circ\circ\circ) \qquad\qquad (\overset{1}{\bullet}\overset{2}{\bullet}\overset{3}{\bullet}) \xrightarrow{f_2} (\circ\circ\circ)(\overset{1}{\bullet}\overset{2}{\bullet}\overset{3}{\bullet})$$

The images of f_1 and f_2 via the functor E_1^ε here turn out to be identities: they map any word to the same copy of that word; e.g. $\texttt{ATG} \mapsto \texttt{ATG}$. Now, since the functor E_1^ε is a $\mathcal{W}^{\text{bij}}$-pedigrad, the image of the obvious exactly 1-distributive cone

$$\begin{tikzcd} & (\bullet\bullet\bullet)(\circ\circ\circ) \\ (\bullet\bullet\bullet)(\bullet\bullet\bullet) \arrow[ur] \arrow[dr] & \\ & (\circ\circ\circ)(\bullet\bullet\bullet) \end{tikzcd} \tag{3.3}$$

via the functor E_1^ε is associated with a concatenation operation of the following form (this is the inverse of the limit adjoint of the image of cone (3.3) via the functor E_1^ε).

$$\mu : E_1^\varepsilon((\bullet\bullet\bullet)(\circ\circ\circ)) \times E_1^\varepsilon((\circ\circ\circ)(\bullet\bullet\bullet)) \longrightarrow E_1^\varepsilon((\bullet\bullet\bullet)(\bullet\bullet\bullet))$$

Here, by concatenation, we mean a map that concatenates words; e.g. the pair of words $(\texttt{ATG}, \texttt{CGG})$ will be sent to the word $\texttt{ATGCGG}$. The post-composition of the morphism μ with the product of morphisms $E_1^\varepsilon(f_1) \times E_1^\varepsilon(f_2)$ then gives rise to a map of the type given below that duplicates any word contained in its domain; e.g. $\texttt{ATG} \mapsto \texttt{ATGATG}$.

$$E_1^\varepsilon((\bullet\bullet\bullet)) \to E_1^\varepsilon((\bullet\bullet\bullet)(\bullet\bullet\bullet))$$

3.5. **Morphisms of pedigrads.** According to section 2.11, a morphism between two pedigrads is a natural transformation between the underlying functors of the pedigrads. The goal of this section is to show how morphisms of pedigrads can be used to model transcription and mutation processes. We shall mainly use Remark 3.31.

Remark 3.31. Let $(\Omega, \preceq)$ be a pre-ordered set and b be an element in Ω. For every map of pointed sets $f : (A, \alpha) \to (B, \beta)$ in $\mathbf{Set}_*$, there is a natural transformation $f_b^* : A_b^\alpha \Rightarrow B_b^\beta$ given by the evaluation of the bifunctor $\mathbf{Set}_*(\mathrm{Tr}_b^*(_), _)$ at f on the second variable.

$$\mathbf{Set}_*(\mathrm{Tr}_b^*(_), f) : \mathbf{Set}_*(\mathrm{Tr}_b^*(_), (A, \alpha)) \Rightarrow \mathbf{Set}_*(\mathrm{Tr}_b^*(_), (B, \beta))$$

The naturality of the resulting transformation $f_b^* : A_b^\alpha \Rightarrow B_b^\beta$ directly follows from the functoriality of the bifunctor in the first variable.

Example 3.32 (Transcription)**.** For convenience, let $(\Omega, \preceq)$ be the Boolean pre-ordered set $\{0 \leq 1\}$ and take b to be $1 \in \Omega$. If we denote by (A, ε) and (B, ε) the pointed sets given by the alphabets $\{\texttt{A}, \texttt{C}, \texttt{G}, \texttt{T}, \varepsilon\}$ and $\{\texttt{A}, \texttt{C}, \texttt{G}, \texttt{U}, \varepsilon\}$, respectively, then we can construct an isomorphism $f : (A, \varepsilon) \to (B, \varepsilon)$ in $\mathbf{Set}_*$ by considering the mappings $\texttt{A} \mapsto \texttt{U}$; $\texttt{T} \mapsto \texttt{A}$; $\texttt{G} \mapsto \texttt{C}$; $\texttt{C} \mapsto \texttt{G}$. The morphism of pedigrads $f_b^* : A_b^\varepsilon \Rightarrow B_b^\varepsilon$ resulting from Remark 3.31 then sends any words of the form given below, on the left, to the corresponding words, on the right.

$$\begin{array}{ccc} A_b^\varepsilon((\bullet\bullet\bullet)(\bullet\bullet\bullet)(\bullet\bullet\bullet)) & \to & B_b^\varepsilon((\bullet\bullet\bullet)(\bullet\bullet\bullet)(\bullet\bullet\bullet)) \\ \texttt{TGTAGTAGC} & \mapsto & \texttt{ACAUCAUCG} \\ \texttt{AAACTTACA} & \mapsto & \texttt{UUUGAAUGU} \end{array}$$

As one can see, this type of transformation models RNA transcription (by sending the thymine nucleobases $(\texttt{T})$ to its uracil anti-nucleobases $(\texttt{U})$).

Remark 3.33 (Coarse mutations). As one can imagine, the type of morphism constructed in Remark 3.31 can be used to represent DNA mutations. However, note that if one retricts oneself to morphisms of pointed sets of the form $(E, \varepsilon) \to (E, \varepsilon)$ where (E, ε) is our usual pointed set $\{\texttt{A}, \texttt{C}, \texttt{G}, \texttt{T}, \varepsilon\}$, then these mutations will obviously be too systematic to be realistic. For instance, see the example given below for the mappings $\texttt{A} \mapsto \varepsilon$; $\texttt{T} \mapsto \texttt{T}$; $\texttt{G} \mapsto \texttt{G}$; $\texttt{C} \mapsto \texttt{A}$, which delete any adenine $(\texttt{A})$ and substitute any cytosine $(\texttt{C})$ with an adenine $(\texttt{A})$.

$$\begin{array}{ccc} E_1^\varepsilon((\bullet\bullet\bullet)(\bullet\bullet\bullet)(\bullet\bullet\bullet)) & \to & E_1^\varepsilon((\bullet\bullet\bullet)(\bullet\bullet\bullet)(\bullet\bullet\bullet)) \\ \texttt{AGCAGTAGC} & \mapsto & \varepsilon\texttt{GA}\varepsilon\texttt{GT}\varepsilon\texttt{GA} \\ \texttt{TAACCTACA} & \mapsto & \texttt{T}\varepsilon\varepsilon\texttt{AAT}\varepsilon\texttt{AA} \end{array}$$

Note that a parameterization of the alphabet could specify the context in which these mutations happen, which would make them more realistic (see Example 3.34).

Finally, note that the definition of E_b^ε forces us to always map the element ε to itself, so that insertion mutations cannot directly be viewed as morphisms of the previous type. In fact, insertion mutations could be encoded as the fibers (i.e. lifts) of a morphism of the form described above. This idea is further discussed in Example 3.34 via the concept of span.

Example 3.34 (Mutations are spans)**.** The present example shows that mutations can be recovered from spans in the underlying category of pedigrads. Let $(\Omega, \preceq)$ be the Boolean pre-ordered set $\{0 \leq 1\}$ and take b to be $1 \in \Omega$. Denote by (E, ε) the pointed set given by the alphabet $\{\texttt{A}, \texttt{C}, \texttt{G}, \texttt{T}, \varepsilon\}$. Since $\mathbf{Set}_*$ is complete, we can form the Cartesian product of (E, ε), which, for convenience, will be denoted as $(A, (\varepsilon, \varepsilon))$, where A is the set of pairs of elements in E; e.g. $(\texttt{A}, \varepsilon)$, $(\texttt{A}, \texttt{T})$, etc. By definition, we are given a span of projection maps

$(A,(\varepsilon,\varepsilon)) \rightrightarrows (E,\varepsilon)$ in $\mathbf{Set}_*$ that forget the second and first components of the pairs in A, respectively. The corresponding span resulting from Remark 3.31 can then be viewed as a binary relation describing all the possible ways a DNA strands can be mutated (the binary relation appears when reading the span of mappings from left to right, as shown below).

$$\begin{array}{ccccc}
E_b^{\varepsilon}((\bullet\bullet\bullet)(\bullet\bullet\bullet)(\bullet\bullet\bullet)) & \leftarrow & A_b^{(\varepsilon,\varepsilon)}((\bullet\bullet\bullet)(\bullet\bullet\bullet)(\bullet\bullet\bullet)) & \rightarrow & E_b^{\varepsilon}((\bullet\bullet\bullet)(\bullet\bullet\bullet)(\bullet\bullet\bullet)) \\
\texttt{TGCAG}\varepsilon\texttt{AG}\varepsilon & \leftarrow\!\shortmid & \binom{\texttt{T}}{\texttt{T}}\binom{\texttt{G}}{\texttt{G}}\binom{\texttt{C}}{\texttt{C}}\binom{\texttt{A}}{\texttt{A}}\binom{\texttt{G}}{\texttt{G}}\binom{\varepsilon}{\texttt{T}}\binom{\texttt{A}}{\texttt{A}}\binom{\texttt{G}}{\texttt{C}}\binom{\varepsilon}{\varepsilon} & \mapsto & \texttt{TGCAGTAC}\varepsilon \\
\texttt{TGCAG}\varepsilon\texttt{AG}\varepsilon & \leftarrow\!\shortmid & \binom{\texttt{T}}{\texttt{A}}\binom{\texttt{G}}{\varepsilon}\binom{\texttt{C}}{\texttt{C}}\binom{\texttt{A}}{\varepsilon}\binom{\texttt{G}}{\texttt{G}}\binom{\varepsilon}{\texttt{A}}\binom{\texttt{A}}{\texttt{A}}\binom{\texttt{G}}{\texttt{G}}\binom{\varepsilon}{\texttt{C}} & \mapsto & \texttt{A}\varepsilon\texttt{C}\varepsilon\texttt{GAAGC}
\end{array}$$

Note that the previous mappings describe the three most common types of mutation, namely deletion, insertion and substitution mutations. We will finally conclude by pointing out that another span $A \rightrightarrows E$ could have been chosen so that the mutations could have well depended on some environmental contexts or other similar conditions.

3.6. **Reversible pedigrads.** Let $(\Omega,\preceq)$ be a pre-ordered set. The goal of this section is to show how one can use a category of pedigrads to model certain biological phenomena, such as the phenomenon of reversion.

Convention 3.35 (Notation)**.** For every non-negative integer n, we will denote by rv_n the function $[n] \to [n]$ sending every element $x \in [n]$ to the interger $n+1-x$ in $[n]$.

Example 3.36. The following picture shows what the map $\mathrm{rv}_6 : [6] \to [6]$ looks like.

$$\begin{array}{ccccccccccc}
1 & \leq & 2 & \leq & 3 & \leq & 4 & \leq & 5 & \leq & 6 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
6 & \geq & 5 & \geq & 4 & \geq & 3 & \geq & 2 & \geq & 1
\end{array}$$

This type of map reverses the order of the finite set on which it is defined.

Remark 3.37. For every non-negative integer n, the function $\mathrm{rv}_n : [n] \to [n]$ is an involution[4].

Definition 3.38 (Reversion operation)**.** For every object $(t,c) : [n_1] \multimap [n_0]$ in $\mathbf{Seg}(\Omega)$, we define the *reversion* of (t,c) as the segment of type $[n_1] \multimap [n_0]$, over Ω, that consists of the order-preserving surjection $\mathrm{rv}_{n_0} \circ t \circ \mathrm{rv}_{n_1} : [n_1] \to [n_0]$ and the function $c \circ \mathrm{rv}_{n_0} : [n_0] \to \Omega$. The reversion of (t,c) will be denoted as $(t,c)^{\dagger}$.

Example 3.39. Let $(\Omega,\preceq)$ be the Boolean pre-ordered set $\{0 \leq 1\}$. The reversion of the segment, over Ω, displayed below, on the left, is shown on the right.

$$(c,t) = \ (\bullet\bullet)(\circ)(\bullet\bullet\bullet)(\bullet)(\circ)(\circ) \qquad \mapsto \qquad (t,c)^{\dagger} = \ (\circ)(\circ)(\bullet)(\bullet\bullet\bullet)(\circ)(\bullet\bullet)$$

Convention 3.40 (Reversion functor)**.** The mapping $(t,c) \mapsto (t,c)^{\dagger}$ defines an obvious endofunctor on $\mathbf{Seg}(\Omega)$ mapping a morphism of segments (f_1,f_0) to the morphism of segments $(f_1^{\dagger},f_0^{\dagger})$ defined by the pair $(\mathrm{rv}_{n_1'} \circ f_1 \circ \mathrm{rv}_{n_1}, \mathrm{rv}_{n_0'} \circ f_0 \circ \mathrm{rv}_{n_0})$, as shown below.

$$\begin{tikzcd}
{[n_1]} \arrow[r, two heads, "t"] \arrow[d, "f_1"'] & {[n_0]} \arrow[d, "f_0"] \\
{[n_1']} \arrow[r, two heads, "t'"] & {[n_0']}
\end{tikzcd}
\qquad \mapsto \qquad
\begin{tikzcd}
{[n_1]} \arrow[r, "\mathrm{rv}_{n_1}"] \arrow[d, "f_1^{\dagger}"'] & {[n_1]} \arrow[r, two heads, "t"] \arrow[d, "f_1"'] & {[n_0]} \arrow[r, "\mathrm{rv}_{n_0}"] \arrow[d, "f_0"] & {[n_0]} \arrow[d, "f_0^{\dagger}"] \\
{[n_1']} \arrow[r, "\mathrm{rv}_{n_1'}"'] & {[n_1']} \arrow[r, two heads, "t'"] & {[n_0']} \arrow[r, "\mathrm{rv}_{n_0'}"'] & {[n_0']}
\end{tikzcd}$$

This endofunctor will be denoted as $\mathrm{Rev} : \mathbf{Seg}(\Omega) \to \mathbf{Seg}(\Omega)$ and called the *reversion functor* on $\mathbf{Seg}(\Omega)$.

Remark 3.41 (Involution). The reversion functor on $\mathbf{Seg}(\Omega)$ is an involution functor and hence an isomorphism of categories.

Definition 3.42 (Reversible chromologies)**.** A chromology (Ω,D) will be said to be *reversible* if the image of every cone in D via the reversion functor is a cone in D.

[4] i.e. a bijection that is its own inverse.

Definition 3.43 (Reversible pedigrads)**.** Let $(\mathcal{C}, \mathcal{W})$ be a logical system. A $\mathcal{W}$-pedigrad X on a reversible chromology (Ω, D) will be said to be *reversible* if it is equipped with an isomorphism $X \Rightarrow X \circ \mathrm{Rev}$ in $[\mathbf{Seg}(\Omega), \mathcal{C}]$.

Example 3.44. Let $(\Omega, \preceq)$ be a the Boolean pre-ordered set $\{0 \leq 1\}$. For every element $b \in \Omega$, the functor $E_b^\varepsilon : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ defines a reversible $\mathcal{W}^{\mathrm{bij}}$-pedigrad on any reversible chromology when it is equipped with the natural transformation that maps every word $\mathtt{X}_1\mathtt{X}_2 \ldots \mathtt{X}_n$ in the domain to its reverse $\mathtt{X}_n \ldots \mathtt{X}_2\mathtt{X}_1$ in the codomain.

$$E_b^\varepsilon(t, c) \to E_b^\varepsilon \circ \mathrm{Rev}(t, c)$$

For instance, we would have the following mappings in the case where E is our usual pointed set $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}, \varepsilon\}$ and b is taken to be $1 \in \Omega$.

$E_1^\varepsilon((\bullet\bullet)(\bullet)(\bullet\bullet\bullet))$	$\to$	$E_1^\varepsilon((\bullet\bullet\bullet)(\bullet)(\bullet\bullet))$		$E_1^\varepsilon((\circ\circ\circ)(\bullet)(\circ\circ))$	$\to$	$E_1^\varepsilon((\circ\circ)(\bullet)(\circ\circ\circ))$
AGTAGC	$\mapsto$	CGATGA		A	$\mapsto$	A
CTTACA	$\mapsto$	ACATTC		C	$\mapsto$	C

Remark 3.45 (Reversible structure). The reversible structure of a reversible $\mathcal{W}^{\mathrm{bij}}$-pedigrad X can be uniquely determined by the cones of the chromology (Ω, D) provided that D contains enough cones. For instance, the components of the reversible structure defined at atomic patches (as shown below) will uniquely determine the entire reversible structure if these atomic patches are part of the cones in D (see Example 3.26)

$$X(\circ\circ\circ\circ\circ\circ\bullet\circ\circ) \to X(\circ\circ\bullet\circ\circ\circ\circ\circ\circ)$$

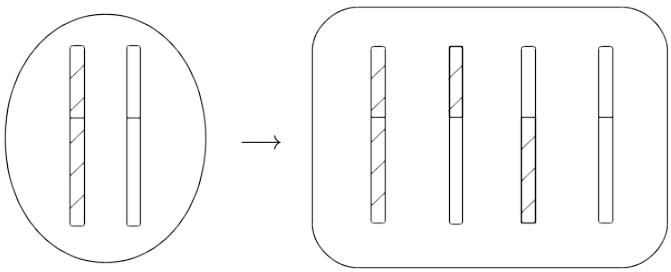
4. Pedigrads in semimodules over semi-rings

In this section, we use our pedigrad $E_b^\varepsilon : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ to generate a new functor in the category of semimodules over the Boolean semi-ring with two elements. We then consider a quotient of this functor to define a pedigrad that models homologous recombination and haplotypes.

A question one might want to ask at this point is why should one consider pedigrads in a category of semimodules specifically? The answer can be given in several points.

1) The first reason is that the tools of semimodule theory allow us to neatly capture the phenomenon of homologous recombination. When we look at the homologous recombination of two segments, say `xxaxxbx` and `xxAxxBx`, we can observe that each one of these will be separated into two parts, say `xxax`, `xxAx` and `xbx`, `xBx`, so that the recombination of `xxaxxbx` and `xxAxxBx` is one of the following re-attachments.

(4.1) `xxaxxbx` ; `xxaxxBx` ; `xxAxxbx` ; `xxAxxBx`



Interestingly, the way the different recombined segments are written in (4.1) is reminiscent of the result of a bilinear operation between the sums $\mathtt{xxax} + \mathtt{xxAx}$ and $\mathtt{xbx} + \mathtt{xBx}$.

$$(\mathtt{xxax} + \mathtt{xxAx}) \otimes (\mathtt{xbx} + \mathtt{xBx}) = \mathtt{xxax} \otimes \mathtt{xbx} + \mathtt{xxax} \otimes \mathtt{xBx} + \mathtt{xxAx} \otimes \mathtt{xbx} + \mathtt{xxAx} \otimes \mathtt{xBx}$$

This suggests us that homologous recombination may belong to an algebraic setting in which the mating operation is given by an addition operation while the possible results of homologous recombination are given by a product operation. The way one passes from the

segments `xxaxxbx` and `xxAxxBx`, or perhaps we should say the sum `xxaxxbx` + `xxAxxBx`, to the sums `xxax` + `xxAx` and `xbx` + `xBx` would be encoded as a functor from $\mathbf{Seg}(\Omega)$ to a category of semimodules.

2) Semimodules over semi-rings are very natural structures in which it is possible to talk about all sorts of interactions without contradicting the noticeable behavior of real processes. Specifically, the reason for which we will use semi-rings instead of rings is that quotients of rings are too destructive. While in semimodules over rings, an equation of the form $x+y=x$ implies the identity $y=0$, in semimodules over semi-rings, an equality $x+y=y$ will inform us that the information contained in x is already contained in y, which is closer to the type of things one may want to say about biological systems.

3) Finally, the Boolean semi-ring with two elements, call it B_2, has the advantage to be suitable for statistical analysis (see section 5). For instance, it is interesting to note that the free B_2-semimodules that is generated over a finite set S is isomorphic to the power set of S equipped with its union operation (see Example 4.2 and section 4.4). This power set, call it $\mathcal{P}(S)$, can then be seen as the set of probable events for a given universe S of outcomes. As a result, the elements of a B_2-semimodules will sometimes be seen as statistical events in the sense of a probability space [14].

4.1. **Semi-rings.** We will call a *semi-ring* [8, page 1] any set R equipped with a commutative monoid structure $(R,+,0)$ and monoid structure $(R,\cdot,1)$ such that $+$ is distributive over $\cdot$ and the operation $\cdot$ annihilates the neutral element 0, that is to say that, for every triple $a,b,c\in R$, the following identities hold.

$$(a+b)\cdot c = a\cdot c + b\cdot c \qquad a\cdot(b+c) = a\cdot b + a\cdot c \qquad a\cdot 0 = 0 = 0\cdot a$$

We shall denote by B_2 the semi-ring consisting of two elements $\{0,1\}$ whose monoid structures are defined according to the following addition and multiplication tables.

$+$	0	1
0	0	1
1	1	1

$\cdot$	0	1
0	0	0
1	0	1

4.2. **Semimodules over semi-rings.** Let $(R,+,\cdot,0,1)$ be a semi-ring. A *semimodule* over R [8, page 149] is a commutative monoid $(M,\oplus,0_M)$ equipped with a function $\odot : R\times M\to M$ satisfying the following axioms:

1) for every $r\in R$ and $x,y\in M$, the identity $r\odot(x\oplus y)=(r\odot x)\oplus(r\odot y)$ holds;
2) for every $r,s\in R$ and $x\in M$, the identity $(r+s)\odot x=(r\odot x)\oplus(s\odot x)$ holds;
3) for every $r,s\in R$ and $x\in M$, the identity $(r\cdot s)\odot x = r\odot(s\odot x)$ holds;
4) for every $r\in R$ and $x\in M$, the identities $1\odot x = x$ and $0\odot x = 0_M = r\odot 0_M$ hold;

In the sequel, the operations $\oplus$ and $\odot$ will be denoted by the same symbols as those used for the semi-ring structure, namely $(_)+(_)$ and $(_)\cdot(_)$.

Example 4.1 (B_2 as a B_2-semimodule)**.** The set B_2 has an obvious semimodule structure over itself where its commutative monoid structure is given by the addition $+ : B_2\times B_2\to B_2$ and the action is given by the multiplication $\cdot : B_2\times B_2\to B_2$.

Example 4.2 (Power sets)**.** Recall that the *power set* of a set S, say $\mathcal{P}(S)$, is the set of all subsets of S. This set can also be viewed as the hom-set $\mathbf{Set}(S,\{0,1\})$, in which a function $S\to\{0,1\}$ can be seen as a subset of S by specifying whether it contains an element or not by sending it to 1 or 0. We illustrate below the different ways that can be used to represent an element in the set $\mathcal{P}(S)$ when $S=\{\mathtt{AG},\mathtt{CG},\mathtt{CA},\mathtt{TA}\}$.

As a subset	As a function	As a formal sum
$f=\{\mathtt{AG},\mathtt{CG},\mathtt{TA}\}$	$f:\begin{bmatrix}\mathtt{AG} & \mapsto & 1 & \mathtt{CA} & \mapsto & 0\\ \mathtt{CG} & \mapsto & 1 & \mathtt{TA} & \mapsto & 1\end{bmatrix}$	$f=\mathtt{AG}+\mathtt{CG}+\mathtt{TA}$

Let us now show that $\mathcal{P}(S)$ has a semimodule structure over B_2. As will be seen, each representation turn out to be useful depending on what one wants to express.

First, the power set $\mathcal{P}(S)$ is a commutative monoid for the union of subsets of S. Note that this monoid operation can also be defined as the pointwise extension of the addition of B_2 on the images of the functions $S \to \{0,1\}$. However, in a calculation, it is better used in terms of formal sums, as shown below.

$$(\mathtt{AG}+\mathtt{CG})+(\mathtt{TA}+\mathtt{CG}) = \mathtt{AG}+\underbrace{(1+1)}_{=1}\cdot\mathtt{CG}+\mathtt{TA} = \mathtt{AG}+\mathtt{CG}+\mathtt{TA}$$

For its part, the action $B_2 \times \mathcal{P}(S) \to \mathcal{P}(S)$ is induced by the action of the multiplication of B_2 on the images of the functions $S \to \{0,1\}$. Interestingly, the resulting B_2-semimodule is equipped with an obvious function $S \to \mathcal{P}(S)$ (mapping $x \in S$ to the singleton $\{x\} \subseteq S$) that can be shown to possess a universal property when S is a finite set (see Remark 4.8).

4.3. **Morphisms of semimodules.** Let R be a semi-ring and M and N be two semimodules over R. A *morphism of semimodules* from M to N is a function $f : M \to N$ such that for every pair $r, s \in R$ and pair $x, y \in M$, the following identity holds.

$$f(r \cdot x + s \cdot y) = r \cdot f(x) + s \cdot f(y) \tag{4.2}$$

Remark 4.3 (Boolean coefficients). In the case of a morphism of B_2-semimodules, it only suffices to verify equation (4.2) when $(r,s) = (1,1)$ and $(r,s) = (0,0)$, namely the equations $f(x+y) = f(x) + f(y)$ and $f(0_M) = 0_N$.

Example 4.4 (Generator)**.** Any morphism $B_2 \to X$ of semimodules over B_2 is of the form $1 \mapsto x$ and $0 \mapsto 0$. This is equivalent to picking an element in the B_2-semimodule X.

Example 4.5. In this example, we will use the notations of Example 4.2. Let S and T denote the set $\{\mathtt{AG}, \mathtt{CG}, \mathtt{CA}, \mathtt{TA}\}$ and $\{\mathtt{TG}, \mathtt{CG}, \mathtt{GA}, \mathtt{TA}\}$, respectively. The function $f : S \to T$ given below, on the left, induces a morphism $f_* : \mathcal{P}(S) \to \mathcal{P}(T)$ of semimodules over B_2 that maps a subset of S to its image via f (see below, on the right).

$$f : \begin{pmatrix} \mathtt{AG} & \mapsto & \mathtt{TG} \\ \mathtt{CG} & \mapsto & \mathtt{CG} \\ \mathtt{CA} & \mapsto & \mathtt{GA} \\ \mathtt{TA} & \mapsto & \mathtt{TA} \end{pmatrix} \quad \rightsquigarrow \quad \begin{array}{lcl} f_*(\mathtt{AG}) & = & \mathtt{TG} \\ f_*(\mathtt{CG}) & = & \mathtt{CG} \\ f_*(\mathtt{AG}+\mathtt{CA}) & = & \mathtt{TG}+\mathtt{GA} \\ f_*(\mathtt{CA}+\mathtt{CG}) & = & \mathtt{GA}+\mathtt{CG} \end{array}$$

Convention 4.6 (Notation)**.** For every semi-ring R, the obvious category whose objects are semimodules over R (also called *R-semimodules*) and whose arrows are morphisms of R-semimodules will be denoted by R-**Mod**.

Remark 4.7 (Power set as a functor). One can check that the mapping rule $S \mapsto \mathcal{P}(S)$ extends to a functor $\mathcal{P} : \mathbf{Set} \to B_2$-**Mod** (see Examples 4.2 & 4.5).

4.4. **Reminder on B_2-semimodules.** A good reference on semirings and semimodules over semirings is [8]. In this section, we recall some useful facts about B_2-semimodules. First of all, recall that B_2-**Mod** is a locally presentable category and therefore admits all limits and colimits. It also follows from this fact that we have an adjunction of the form (4.3), where the left adjoint F maps a set S to the free B_2-semimodules generated over S (see Remark 4.8) while the right adjoint U is the obvious forgetful functor.

$$\mathbf{Set} \underset{U}{\overset{F}{\underset{\longleftarrow}{\overset{\longrightarrow}{\perp}}}} B_2\text{-}\mathbf{Mod} \tag{4.3}$$

Remark 4.8 (Free semimodules). For every set S, the B_2-semimodule $F(S)$ is equivalently:

1) the set of finite subsets of S;
2) the set of functions $f : S \to \{0,1\}$ for which the fiber $f^{-1}(1)$ is finite;
3) the set of formal sums of finite collections of distinct elements in S.

The unit of adjunction (4.3) is given, in the first case, by the function $x \mapsto \{x\}$ and, in the second case, by the function $x \mapsto \delta_x$ where δ_x maps x to 1 and any other element to 0.

Convention 4.9 (Support)**.** For every set S and element $x \in F(S)$, we will speak of the *support* of x to refer to the set representation of x (see item 1 of Remark 4.8). This set will be denoted as $\mathrm{Supp}(x)$.

Example 4.10 (Support)**.** With respect to the notation of Example 4.2, the support of the formal sum $f = \mathtt{AG} + \mathtt{CG} + \mathtt{TA}$ is the set $\{\mathtt{AG}, \mathtt{CG}, \mathtt{TA}\}$.

Remark 4.11 (Support)*.* For every set S, if we see an element $x \in F(S)$ as a function $S \to \{0, 1\}$, then the set $\mathrm{Supp}(x)$ is equal to the fiber $x^{-1}(1)$.

Definition 4.12 (Sub-element)**.** Let S be a set. An element $y \in F(S)$ will be said to be a *sub-element* of another element $x \in F(S)$ if the support of y is included in the support of x. We will then write $y \leq x$.

Remark 4.13 (Reformulation)*.* From the point of view of the functions $S \to \{0, 1\}$, the relation $\leq$ given in Definition 4.12 can be seen as the pointwise pre-order induced by the Boolean pre-ordered set $\{0 \leq 1\}$ on the following product set (as in Example 2.4).

$$\mathbf{Set}(S, \{0, 1\}) \cong \prod_{s \in S} \{0, 1\}$$

Example 4.14. Let S denote the set $\{\mathtt{AG}, \mathtt{CG}, \mathtt{CA}, \mathtt{TA}\}$. If one wants to think of the relation $\leq$ on $F(S)$ in terms of formal sums, then one deals with inequalities as follows.

$$\mathtt{AG} + \mathtt{CA} \leq \mathtt{AG} + \mathtt{CG} + \mathtt{CA}$$

Proposition 4.15. *The binary relation $\leq$ is a partial order.*

Proof. Because the inclusion of sets is a partial order. □

4.5. **Reminder on monomorphisms and epimorphisms.** In this section, we recall a few facts on how to detect epimorphisms and monomorphisms. The content of this section is therefore about certain concepts of category theory (see below and [27, def. 7.2.1.4]) and not biology.

First, recall that a monomorphism in a category $\mathcal{C}$ is an arrow $m : A \to B$ such that for every pair of parallel arrows $f, g : X \rightrightarrows A$ for which the equation $m \circ f = m \circ g$ holds, the two arrows f and g must be equal.

Proposition 4.16. *If $\mathcal{C}$ has pullbacks, then an arrow $m : A \to B$ is a monomorphism in $\mathcal{C}$ if the pullback $p_1, p_2 : P \rightrightarrows A$ of two copies of m is such that p_1 equals p_2.*

Proof. For every pair of parallel arrows $f, g : X \rightrightarrows A$ for which the equation $m \circ f = m \circ g$ holds, the universality of the pullback P gives an arrow $h : X \to P$ for which the identities $f = p_1 \circ h$ and $g = p_2 \circ h$ hold. Because $p_1 = p_2$, we have $f = g$ and the statement follows. □

Now, recall that an epimorphism in a category $\mathcal{C}$ is an arrow $e : A \to B$ such that for every pair of parallel arrows $f, g : X \rightrightarrows A$ for which the equation $f \circ e = g \circ e$ holds, the two arrows f and g must be equal.

Definition 4.17 (Orthogonality)**.** A morphism $f : X \to Y$ will be said to be *orthogonal* to an object I if for every arrow $i : I \to Y$, there exists a dashed arrow (called the *lift*) making the diagram given below commute.

$$\begin{tikzcd} & X \arrow[d, "e"] \\ I \arrow[ur, dashed] \arrow[r, "i"'] & Y \end{tikzcd}$$

Proposition 4.18. *Every morphism in B_2-**Mod** that is orthogonal with respect to the B_2-semimodule B_2 (Example 4.1) is an epimorphism.*

Proof. We can use Example 4.4 to see that if an arrow $e : X \to Y$ is orthogonal with respect to the B_2-semimodule B_2, then it is surjective: for every $y \in Y$, there exists $x \in X$ for which the identity $e(x) = y$ holds. We can directly check that surjective morphisms in B_2-**Mod** are epimorphisms in B_2-**Mod**. □

Example 4.19 (Coequalizers)**.** Let us show that coequalizer maps [27, def. 3.3.3.1] in B_2-**Mod** are orthogonal to the B_2-semimodule B_2. Let $f, g : X \rightrightarrows Y$ be a pair of morphisms in B_2-**Mod**. Since B_2-**Mod** is cocomplete, we can form its coequalizer $e : Y \to Q$. We can then show, from [8, Example 15.1], that this map is a quotient map. To show this, define the equivalence relation R containing the pair of elements m and m' in Y such that for every object Z and morphism $h : Y \to Z$ in B_2-**Mod** for which the equation $h \circ f = h \circ g$ holds, the relation $h(m) = h(m')$ holds too. We can verify that R defines a congruence [8] in B_2-**Mod** (i.e. an equivalence relation living in B_2-**Mod**). The form of the definition of R is such that all the assumptions of [8, Example 15.1] can be easily verified. It then follows that the map $e : Y \to Q$ can be identified as a quotient map $e_R : M \to M/R$ (see [8, Page 163]) so that there is a bijection making the following diagram commute.

$$\begin{tikzcd} & B_2\text{-}\mathbf{Mod}(B_2, Y) \arrow[dl, "e"'] \arrow[dr, "e_R"] & \\ B_2\text{-}\mathbf{Mod}(B_2, Q) & \cong & B_2\text{-}\mathbf{Mod}(B_2, Y)/R \end{tikzcd}$$

This bijection implies that for every arrow $i : B_2 \to Q$, there exists an arrow $h : B_2 \to Y$ for which the identity $e \circ h = i$ holds. In other words, the coequalizer map e is orthogonal to the B_2-semimodule B_2.

4.6. **Wide spans.** Let $\mathcal{C}$ be a category. We will speak of a *wide span* in $\mathcal{C}$ to refer to a pair $(k, \mathbf{S})$ where k is a positive integer and $\mathbf{S}$ is a $[k]$-indexed collection of arrows in $\mathcal{C}$ whose domains are all equal (see below).

$$\begin{tikzcd} & & S \arrow[dll] \arrow[dl] \arrow[dr] \arrow[drr] & & \\ S_1 & S_2 & \dots & S_{k-1} & S_k \end{tikzcd}$$

Later on, a wide span $(k, \mathbf{S})$ will often be denoted as $\mathbf{S}$ only.

Example 4.20. See Example 4.28 for an example of wide span related to pedigrads.

Remark 4.21 (Wide spans are cones)*.* A wide span is a cone defined over a finite discrete small category whose objects are equipped with a total order.

Definition 4.22 (Cardinality)**.** For every wide span $\mathbf{S} = \{S \to S_i\}_{i \in [k]}$ in $\mathcal{C}$, the integer k and ordered set $[k]$ will be denoted as $|\mathbf{S}|$ and $[\mathbf{S}]$, respectively. Both data will be referred to as the *cardinality* of $\mathbf{S}$ (one being an integer, the other being a set).

Convention 4.23 (Product of a wide span)**.** Suppose $\mathcal{C}$ has products. For every wide span $\mathbf{S} = \{S \to S_i\}_{i \in [k]}$ in $\mathcal{C}$, we will denote by $\mathbf{S}^\times$ the product object $S_1 \times \cdots \times S_n$.

Remark 4.24 (Canonical arrows)*.* Every wide span $\mathbf{S} = \{S \to S_i\}_{i \in [k]}$ in a category $\mathcal{C}$ that has products is equipped with an obvious arrow $S \to \mathbf{S}^\times$ (the limit adjoint of the underlying cone) and an obvious projection map $\mathbf{S}^\times \to S_i$ whose composition is equal to the arrow $S \to S_i$ for every $i \in [k]$.

Definition 4.25 (Cartesian)**.** A wide span $\mathbf{S} = \{S \to S_i\}_{i \in [k]}$ in a category $\mathcal{C}$ that has products $\mathcal{C}$ will be said to be *Cartesian* if the canonical arrow $S \to \mathbf{S}^\times$ is an isomorphism.

4.7. **Finite chromologies.** We shall speak of a *finite chromology* to refer to a chromology (Ω, D) whose sets $D[n]$ are finite sets of wide spans, for every non-negative integer n.

Example 4.26. The cones given in Examples 3.18 & 3.19 are suitable for defining a finite chromology. These cones are defined over a finite discrete small category consisting of three

objects. An example of non-suitable cone is given below, where the small category on which it is defined is a cospan $A = \{\cdot \to \cdot \leftarrow \cdot\}$ and is therefore not discrete.

$$\begin{tikzcd}[column sep=small] & (\bullet\bullet\bullet\bullet)(\circ\circ\circ) \arrow[dr] & \\ (\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet) \arrow[ur] \arrow[dr] & & (\circ\circ\circ\circ)(\circ\circ\circ) \\ & (\circ\circ\circ\circ)(\bullet\bullet\bullet) \arrow[ur] & \end{tikzcd} \qquad \text{(non-suitable)}$$

4.8. **Recombination cones.** In this section, we shall let (Ω, D) denote a finite chromology. The goal of this section is to use the cones of D to associate every segments over Ω with wide span of sets that describe the range of recombination operations related to this object from the point of view of a given pedigrad in **Set** (see section 4.9).

Definition 4.27 (Recombination cones)**.** Let X be a functor $\mathbf{Seg}(\Omega) \to \mathbf{Set}$. For every wide span $\rho : \Delta_A(\tau) \Rightarrow \theta$ in D, the wide span of sets that is the image of ρ via X will be called the *ρ-recombination cone* of X and denoted by $X(\rho)$.

Example 4.28 (Recombination cones)**.** Let Ω be the Boolean pre-ordered set $\{0 \leq 1\}$ and let (E, ε) be our usual pointed set $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}, \varepsilon\}$. We will take b to be equal to 1 and ρ to be the exactly 1-distributive cone given in Example 3.18 (see below); the small discrete category A on which ρ is defined will be taken to be equal to the finite set $\{a_1, a_2, a_3\}$.

$$\begin{tikzcd}[column sep=small] & & (\circ\circ\circ)(\circ\circ)(\bullet\bullet\bullet\bullet)(\circ\circ\circ\circ\circ)(\circ\circ\circ)(\circ) \quad \theta(a_1) \\ \rho : & \underbrace{(\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ)}_{\tau} \arrow[ur] \arrow[r] \arrow[dr] & (\circ\circ\circ)(\circ\circ)(\circ\circ\circ\circ)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) \quad \theta(a_2) \\ & & (\bullet\bullet\bullet)(\circ\circ)(\circ\circ\circ\circ)(\circ\circ\circ\circ\circ)(\circ\circ\circ)(\circ) \quad \theta(a_3) \end{tikzcd}$$

The ρ-recombination cone of E_b^ε is a wide span consisting of three arrows

$$E_b^\varepsilon(\tau) \to E_b^\varepsilon(\theta(a_1)), \quad E_b^\varepsilon(\tau) \to E_b^\varepsilon(\theta(a_2)), \quad E_b^\varepsilon(\tau) \to E_b^\varepsilon(\theta(a_3))$$

whose respective codomains are, according to Remark 3.12, of the form given below, in the rightmost column of the displayed table (the left columns are only given for the sake of exposition) and whose domains $E_b^\varepsilon(\tau)$ are isomorphic to $E^{\times 12}$.

i	$\mathrm{Tr}_1(\theta(a_i))$	$\rightsquigarrow$	$E_b^\varepsilon(\theta(a_i))$	
1	$\{6, 7, 8, 9\}$	$\rightsquigarrow$	$\mathbf{Set}(\{6, \ldots, 9\}, E)$	$\cong E^{\times 4}$
2	$\{10, 11, 12, 13, 14\}$	$\rightsquigarrow$	$\mathbf{Set}(\{10, \ldots, 14\}, E)$	$\cong E^{\times 5}$
3	$\{1, 2, 3\}$	$\rightsquigarrow$	$\mathbf{Set}(\{1, 2, 3\}, E)$	$\cong E^{\times 3}$

By Proposition 3.28, the canonical arrow $E_b^\varepsilon(\tau) \to E_b^\varepsilon(\rho)^\times$ must be a bijection, which implies that the wide span $E_b^\varepsilon(\rho)$ is Cartesian. From the point of view of the isomorphisms given in the previous table, the bijection $E_b^\varepsilon(\tau) \to E_b^\varepsilon(\rho)^\times$ is of the form

$$E^{\times 12} \to E^{\times 4} \times E^{\times 5} \times E^{\times 3}$$

and sends the words given below, on the left, (to which parenthesis have been added to facilitate the recognition of the construction) to the corresponding tuples given on the right.

$$\begin{array}{ccc} (\mathtt{TAG})(\mathtt{ACGA})(\mathtt{CG}\varepsilon\mathtt{TT}) & \mapsto & (\mathtt{ACGA}, \mathtt{CG}\varepsilon\mathtt{TT}, \mathtt{TAG}) \\ (\varepsilon\mathtt{CA})(\mathtt{GGTA})(\mathtt{CCTAT}) & \mapsto & (\mathtt{GGTA}, \mathtt{CCTAT}, \varepsilon\mathtt{CA}) \\ (\mathtt{C}\varepsilon\varepsilon)(\mathtt{GGCC})(\mathtt{TAGTT}) & \mapsto & (\mathtt{GGCC}, \mathtt{TAGTT}, \mathtt{C}\varepsilon\varepsilon) \\ & \text{etc.} & \end{array}$$

4.9. **Recombination congruences.** The goal of this section is to define congruences that will allow us to identify two populations sharing the same haplotype. These congruences will be defined with respect to wide spans as given in section 4.8. In this section, the functor $F : \mathbf{Set} \to B_2\text{-}\mathbf{Mod}$ refers the left adjoint of (4.3).

Convention 4.29 (Wide spans of semimodules)**.** For every wide span $\mathbf{S} = \{S \to S_k\}_{k \in [\mathbf{S}]}$ in **Set**, we will denote by $F(\mathbf{S})$ the corresponding wide span $\{F(S) \to F(S_k)\}_{k \in [\mathbf{S}]}$ in B_2-**Mod**.

Convention 4.30. For every wide span $\mathbf{S} = \{S \to S_k\}_{k\in[\mathbf{S}]}$ in **Set**, we will denote by $\pi_{\mathbf{S}}$ the canonical morphism $F(S) \to F(\mathbf{S})^\times$ in B_2-**Mod** associated with the wide span $F(\mathbf{S})$ in B_2-**Mod**.

Definition 4.31 (Recombination congruences)**.** For every wide span $\mathbf{S} = \{S \to S_k\}_{k\in[\mathbf{S}]}$ in **Set**, the pullback $G(\mathbf{S}) \rightrightarrows F(S)$ defined below will be called the *recombination congruence* of $\mathbf{S}$.

$$\begin{tikzcd} G(\mathbf{S}) \arrow[r, "\mathrm{prj}_1"] \arrow[d, "\mathrm{prj}_2"'] \arrow[dr, phantom, "\lrcorner"] & F(S) \arrow[d, "\pi_{\mathbf{S}}"] \\ F(S) \arrow[r, "\pi_{\mathbf{S}}"'] & F(\mathbf{S})^\times \end{tikzcd}$$

Remark 4.32. If we let $\mathbf{S}$ be a Cartesian wide span in **Set**, then the pullback $G(\mathbf{S}) \rightrightarrows F(S)$ is also defined over the B_2-semimodule $F(\mathbf{S}^\times)$ via the isomorphism $F(S) \to F(\mathbf{S}^\times)$ since the following diagram commutes by universality of $F(\mathbf{S})^\times$.

$$\begin{tikzcd} F(S) \arrow[r, "\cong"] \arrow[rr, bend left, "\pi_{\mathbf{S}}"] & F(\mathbf{S}^\times) \arrow[r] & F(\mathbf{S})^\times \end{tikzcd}$$

Example 4.33 (Homologous recombination)**.** For convenience, we shall let $\mathbf{S}$ denote the Cartesian wide span $E_b^\varepsilon(\rho)$ given in Example 4.28. The associated recombination congruence is of the form $G(\mathbf{S}) \rightrightarrows FE_b^\varepsilon(\tau)$. As explained in Remark 4.32, this recombination congruence can also be viewed as a congruence of the form $G(\mathbf{S}) \rightrightarrows F(\mathbf{S}^\times)$, which will be much easier for the eye (see below). To see what the elements of this congruence look like, take x and y to be the sums given below in the top and bottom rows, respectively.

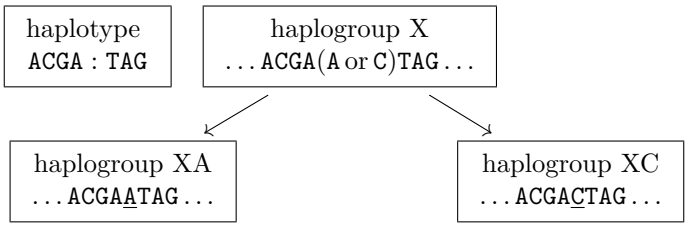
	seen in $F(\mathbf{S}^\times)$	$\cong$	seen in $FE_b^\varepsilon(\tau)$
x	$(\mathtt{ACGA}, \mathtt{CG}\varepsilon\mathtt{TT}, \mathtt{TAG}) + (\mathtt{GGTA}, \mathtt{CCTAT}, \varepsilon\mathtt{CA})$		$\mathtt{TAGACGACG}\varepsilon\mathtt{TT} + \varepsilon\mathtt{CAGGTACCTAT}$
y	$(\mathtt{GGTA}, \mathtt{CG}\varepsilon\mathtt{TT}, \varepsilon\mathtt{CA}) + (\mathtt{ACGA}, \mathtt{CCTAT}, \mathtt{TAG})$		$\varepsilon\mathtt{CAACGACCTAT} + \mathtt{TAGGGTACG}\varepsilon\mathtt{TT}$

We can check that the images $\pi_{\mathbf{S}}(x)$ and $\pi_{\mathbf{S}}(y)$ are equal to the following tuple in $F(\mathbf{S})^\times$.

$$\Big(\mathtt{ACGA} + \mathtt{GGTA}\,,\ \mathtt{CG}\varepsilon\mathtt{TT} + \mathtt{CCTAT}\,,\ \mathtt{TAG} + \varepsilon\mathtt{CA}\Big)$$

In other words, the pair (x, y) belongs to the set $G(\mathbf{S})$. Here, the intuition is that the set $G(\mathbf{S})$ contains pairs of elements that are the same up to homologous recombination with respect to the topology specified by the cone ρ.

Example 4.34 (Haplogroups and haplotypes)**.** In genetics, a *haplotype* is a given set of genes or DNA strands, say `ACGA` and `TAG`, for particular loci on a chromosome while a *haplogroup* for this haplotype can be viewed as a group of DNA segments sharing these strands at the specified locations whereas the other locations may contain single-nucleotide polymorphisms (SNP), which are one-nucleobase long mutations that are noticeable within a non-negligible percentage of the population (see [10, 6]).



From the point of view of Example 4.33, the haplogroup X given in the previous picture can be viewed as an element in the fiber of the map $\pi_{\mathbf{S}} : F(S) \to F(\mathbf{S})^\times$ above the following tuple in $F(\mathbf{S})^\times$.

$$(\mathtt{ACGA}, \mathtt{A} + \mathtt{C}, \mathtt{TAG})$$

In this paper, we shall therefore speak of a 'haplotype' to refer to an element in the B_2-module $F(\mathbf{S})^\times$ while we shall speak of a 'haplogroup' for this haplotype to refer to an element in the fiber of the map $\pi_{\mathbf{S}} : F(S) \to F(\mathbf{S})^\times$ above this element. In Example 4.33, the two elements x and y could be viewed as two haplogroups of the same haplotype $(\mathtt{ACGA} + \mathtt{GGTA}\,,\, \mathtt{CG}\varepsilon\mathtt{TT} + \mathtt{CCTAT}\,,\, \mathtt{TAG} + \varepsilon\mathtt{CA})$. Here, genetic polymorphism is no longer unary.

Remark 4.35 (Haplotypes *versus* recombination). While the concept of haplotype is possible because of the absence of recombination in the transfer of the mtDNA [10], it makes perfect sense to see it arise from an object that is meant to model recombination. As illustrated in Example 4.34, this absence of recombination is modelled by singleton elements, which charactarize the haplotype, whereas the non-singleton elements (i.e. the sums) encode the genetic polymorphism associated with a population. In general, the classification of haplogroups relies on single-nucleotide polymorphisms (SNP) while the polymorphism pertaining to rather-long DNA intervals is usually the type of variations that is studied from the point of view of homologous recombination [9].

Remark 4.36 (Congruence). For every wide span $\mathbf{S}$ in $\mathbf{Set}$, the recombination congruence $G(\mathbf{S}) \rightrightarrows F(S)$ defines an actual congruence (i.e. an equivalence relation) in B_2-$\mathbf{Mod}$. Checking that the pullback $G(\mathbf{S})$ is a sub-object of the product $F(S) \times F(S)$ is straightforward. The reflexivity, symmetry and transitivity axioms follow after noticing that the pullback is defined over a cospan made of two copies of the same arrow (see Definition 4.31).

Convention 4.37 (Notation)**.** Let $\mathbf{S}$ be a wide span in $\mathbf{Set}$. For every pair of elements $x, y \in F(S)$, we shall write $x \sim_{\mathbf{S}} y$ to mean that the pair (x, y) belongs to $G(\mathbf{S})$.

Example 4.38 (Haplotypes)**.** By definition of $G(\mathbf{S})$, the two elements x and y given in Example 4.33 are also equivalent to the sum $x + y$ displayed below.

$$(\mathtt{ACGA}, \mathtt{CG}\varepsilon\mathtt{TT}, \mathtt{TAG}) + (\mathtt{GGTA}, \mathtt{CCTAT}, \varepsilon\mathtt{CA}) + (\mathtt{GGTA}, \mathtt{CG}\varepsilon\mathtt{TT}, \varepsilon\mathtt{CA}) + (\mathtt{ACGA}, \mathtt{CCTAT}, \mathtt{TAG})$$

To see this, observe that we have the relations $x \sim_{\mathbf{S}} y$ and $x \sim_{\mathbf{S}} x$. The fact that $G(\mathbf{S})$ is a B_2-semimodule then gives us the relation $x + x \sim_{\mathbf{S}} x + y$. Since the equation $x + x = x$ holds, we obtain the relation $x \sim_{\mathbf{S}} x + y$. Similarly, we can show that the relation $y \sim_{\mathbf{S}} x + y$ holds too. Intuitively, this means that if x and y are two populations of the same haplotype (i.e. $x \sim_{\mathbf{S}} y$), then the union of these is of the same haplotype.

The argument of Example 4.38 could be formalized in terms of a partial order on the so-called haplogroups so that the sum of all the elements contained in the equivalence class of a recombination congruence is the maximum haplogroup, which could be viewed as the haplotype itself. The rest of this section shows that this representative is obtained from a product operation (see Convention 4.39) on the local patches of every haploroup contained in the equivalence class.

Convention 4.39 (Notation)**.** For every wide span $\mathbf{S}$, we will denote by $\beta_{\mathbf{S}}$ the function $U(F(\mathbf{S})^\times) \to UF(\mathbf{S}^\times)$ that maps a tuple $(x_1, x_2, \ldots, x_{|\mathbf{S}|})$ in $F(\mathbf{S})^\times$ to the element of $F(\mathbf{S}^\times)$ represented by the finite subset $\mathrm{Supp}(x_1) \times \mathrm{Supp}(x_2) \times \cdots \times \mathrm{Supp}(x_{|\mathbf{S}|}) \subseteq \mathbf{S}^\times$.

Example 4.40 (Sequel of Example 4.33)**.** Let $\mathbf{S}$ be the Cartesian wide span $E_b^\varepsilon(\rho)$ given in Example 4.28. If we take $x \in F(\mathbf{S}^\times)$ to be the element considered in Example 4.33, then the element

$$\beta_{\mathbf{S}}(\pi_{\mathbf{S}}(x)) = \{\mathtt{ACGA}, \mathtt{GGTA}\} \times \{\mathtt{CG}\varepsilon\mathtt{TT}, \mathtt{CCTAT}\} \times \{\mathtt{TAG}, \varepsilon\mathtt{CA}\}$$

is represented in $F(\mathbf{S}^\times)$ by the following subset of $\mathbf{S}^\times$.

$$\left\{ \begin{array}{llll} (\mathtt{ACGA}, \mathtt{CG}\varepsilon\mathtt{TT}, \mathtt{TAG}) & (\mathtt{ACGA}, \mathtt{CG}\varepsilon\mathtt{TT}, \varepsilon\mathtt{CA}) & (\mathtt{ACGA}, \mathtt{CCTAT}, \mathtt{TAG}) & (\mathtt{ACGA}, \mathtt{CCTAT}, \varepsilon\mathtt{CA}) \\ (\mathtt{GGTA}, \mathtt{CG}\varepsilon\mathtt{TT}, \mathtt{TAG}) & (\mathtt{GGTA}, \mathtt{CG}\varepsilon\mathtt{TT}, \varepsilon\mathtt{CA}) & (\mathtt{GGTA}, \mathtt{CCTAT}, \mathtt{TAG}) & (\mathtt{GGTA}, \mathtt{CCTAT}, \varepsilon\mathtt{CA}) \end{array} \right\}$$

Interestingly, the image of this element via the canonical arrow $F(\mathbf{S}^\times) \to F(\mathbf{S})^\times$, which can here be viewed as the map $\pi_{\mathbf{S}} : FE_b^\varepsilon(\tau) \to F(\mathbf{S})^\times$ by Remark 4.32, is equal to the following tuple.

$$\Big(\mathtt{ACGA} + \mathtt{GGTA}\,,\, \mathtt{CG}\varepsilon\mathtt{TT} + \mathtt{CCTAT}\,,\, \mathtt{TAG} + \varepsilon\mathtt{CA} \Big)$$

By Example 4.33, this tuple is equal to the element $\pi_{\mathbf{S}}(x)$, which implies that the identity $\pi_{\mathbf{S}}(x) = \pi_{\mathbf{S}}\beta_{\mathbf{S}}(\pi_{\mathbf{S}}(x))$ holds. In terms of haplogroups, this means that the two populations represented by x and $\beta_{\mathbf{S}}(\pi_{\mathbf{S}}(x))$ have the same haplotype (i.e. $x \sim_{\mathbf{S}} \beta_{\mathbf{S}}(\pi_{\mathbf{S}}(x))$). Proposition 4.43, given below, explains why such a relation holds in a greater generality.

Convention 4.41 (Notation)**.** For every wide span $\mathbf{S}$ in $\mathbf{Set}$, we will denote by $\perp$ the function $U(F(\mathbf{S})^\times) \to U(F(\mathbf{S})^\times)$ that maps a tuple $(x_1, x_2, \ldots, x_{|\mathbf{S}|}) \in F(\mathbf{S})^\times$ to 0 if the tuple contains a zero component and that maps the tuple to itself otherwise.

Remark 4.42. It follows from Conventions 4.41 & 4.30 that the following diagram commutes.

$$\begin{tikzcd} UF(S) \arrow[r, "U\pi_{\mathbf{S}}"] \arrow[rr, bend left, "U\pi_{\mathbf{S}}"] & U(F(\mathbf{S})^\times) \arrow[r, "\perp"] & U(F(\mathbf{S})^\times) \end{tikzcd}$$

This comes from the fact that the only element of $F(S)$ whose image, in $F(\mathbf{S})^\times$, contains a zero component is the zero element of $F(S)$.

Proposition 4.43. *Let $\mathbf{S}$ be a wide span of sets. The following diagram commutes in $\mathbf{Set}$, where the rightmost vertical arrow is the image of the obvious arrow $F(\mathbf{S}^\times) \to F(\mathbf{S})^\times$ via the functor U.*

$$\begin{tikzcd} U(F(\mathbf{S})^\times) \arrow[r, "\beta_{\mathbf{S}}"] \arrow[dr, "\perp"'] & UF(\mathbf{S}^\times) \arrow[d] \\ & U(F(\mathbf{S})^\times) \end{tikzcd}$$

Proof. Suppose that $\mathbf{S}$ is of the form $\{S \to S_i\}_{i\in[\mathbf{S}]}$ and take $\tilde{x} = (x_1, x_2, \ldots, x_{|\mathbf{S}|})$ to be a tuple of $F(\mathbf{S})^\times$. If the components of $\tilde{x}$ are all non-zero elements, the arrow $F(\mathbf{S}^\times) \to F(S_i)$ sends the element $\beta_{\mathbf{S}}(\tilde{x})$ represented by the subset $\mathrm{Supp}(x_1) \times \mathrm{Supp}(x_2) \times \cdots \times \mathrm{Supp}(x_{|\mathbf{S}|}) \subseteq \mathbf{S}^\times$ to the element represented by the subset $\mathrm{Supp}(x_i) \subseteq S_i$. Because this element can be identified as x_i itself, this means that the arrow $F(\mathbf{S}^\times) \to F(\mathbf{S})^\times$ sends $\beta_{\mathbf{S}}(\tilde{x})$ to the tuple $\tilde{x}$. If there exists $k \in [\mathbf{S}]$ such that x_k is zero, then the definition of a cartesian product on an empty set implies that $\beta_{\mathbf{S}}(\tilde{x})$ is zero, so that its image via the arrow $F(\mathbf{S}^\times) \to F(S_i)$ is also zero. In this case, the arrow $F(\mathbf{S}^\times) \to F(\mathbf{S})^\times$ sends $\beta_{\mathbf{S}}(\tilde{x})$ to zero. □

The reader can easily verify the following statement in the case of Examples 4.33 & 4.40.

Proposition 4.44. *For every wide span $\mathbf{S}$ of sets and element $x \in F(\mathbf{S}^\times)$, the inclusion $\mathrm{Supp}(x) \subseteq \beta_{\mathbf{S}}(\pi_{\mathbf{S}}(x))$ holds.*

Proof. Suppose that $\mathbf{S}$ is of the form $\{S \to S_i\}_{i\in[k]}$. For any element $s \in \mathbf{S}^\times$, we will use the notation $s(j)$ to denote the image of s via the j-th projection $\mathbf{S}^\times \to S_j$. Without loss of generality, every element $x \in F(\mathbf{S}^\times)$ can be supposed to be of the form $\sum_{k=1}^{n} s_k$ where $s_k \in \mathbf{S}^\times$. Now, observe that, for every $k \in \{1, \ldots, n\}$, the tuple $(s_k(1), \ldots, s_k(|\mathbf{S}|))$, which can be viewed as the element $s_k \in \mathbf{S}^\times$ itself, must belong to the set

$$\beta_{\mathbf{S}}(\pi_{\mathbf{S}}(x)) = \{s_k(1) \mid k \in [n]\} \times \{s_k(2) \mid k \in [n]\} \times \cdots \times \{s_k(|\mathbf{S}|) \mid k \in [n]\},$$

so that the inclusion $\mathrm{Supp}(x) \subseteq \beta_{\mathbf{S}}(\pi_{\mathbf{S}}(x))$ holds. □

The following definition recalls what a maximum, in a partially ordered set, is.

Definition 4.45 (Maximum)**.** Let $(G, \leq)$ be a partially ordered set. An element $u \in G$ will be called a *maximum* if for all $x \in S$, the relation $x \leq u$ holds;

Remark 4.46 (Maxima are unique). It follows from the anti-symmetry axiom of partial orders that if a partially ordered set $(G, \leq)$ admits a maximum, then there is no other maximum. In other words, the maximum of G is unique.

Theorem 4.47 (Haplotypes)**.** *Let* $\mathbf{S}$ *be a Cartesian wide span in* $\mathbf{Set}$*. Every equivalence class of* $G(\mathbf{S})$ *admits a maximum for the partial order of Proposition 4.15. The maximum in the equivalence class of an element* x *is given by the element in* $F(S)$ *that is represented by the subset* $\beta_{\mathbf{S}}\pi_{\mathbf{S}}(x) \subseteq \mathbf{S}^{\times}$ *in* $F(\mathbf{S}^{\times})$ *throught the isomorphism* $F(S) \to F(\mathbf{S}^{\times})$*.*

Proof. Proposition 4.43 and Remark 4.42 implies the equation $\pi_{\mathbf{S}}(x) = \pi_{\mathbf{S}}(\beta_{\mathbf{S}}\pi_{\mathbf{S}}(x))$ and hence the relation $x \sim_{\mathbf{S}} \beta_{\mathbf{S}}\pi_{\mathbf{S}}(x)$. The fact that $\beta_{\mathbf{S}}\pi_{\mathbf{S}}(x)$ is a maximum follows from Definition 4.12 and Proposition 4.44. □

Example 4.48 (Haplogroups)**.** From the point of view of Examples 4.40 & 4.38, the maximum element of Proposition 4.47 is what we could call the *saturated* haplogroup of x (or y), that is to say the element represented by the following sum in $F(\mathbf{S}^{\times})$.

$$(\mathtt{ACGA}, \mathtt{CG}\varepsilon\mathtt{TT}, \mathtt{TAG}) + (\mathtt{ACGA}, \mathtt{CG}\varepsilon\mathtt{TT}, \varepsilon\mathtt{CA}) + (\mathtt{ACGA}, \mathtt{CCTAT}, \mathtt{TAG}) + (\mathtt{ACGA}, \mathtt{CCTAT}, \varepsilon\mathtt{CA}) +$$
$$(\mathtt{GGTA}, \mathtt{CG}\varepsilon\mathtt{TT}, \mathtt{TAG}) + (\mathtt{GGTA}, \mathtt{CG}\varepsilon\mathtt{TT}, \varepsilon\mathtt{CA}) + (\mathtt{GGTA}, \mathtt{CCTAT}, \mathtt{TAG}) + (\mathtt{GGTA}, \mathtt{CCTAT}, \varepsilon\mathtt{CA})$$

Remark 4.49. If the wide span $\mathbf{S}$ of Proposition 4.47 were not Cartesian, then nothing would ensure us that the saturated haplogroup of Example 4.48 can be lifted to $FE_b^{\varepsilon}(\tau)$. The maximum representative of the equivalence class would then be missing certain elements of the saturated haplogroup so that it could not be seen as coming from a product operation.

4.10. **Recombination semimodules.** Let $(\Omega, \preceq)$ be a pre-ordered set. We shall let (Ω, D) denote a finite chromology and X be a functor $\mathbf{Seg}(\Omega) \to \mathbf{Set}$.

Recall that, on the one hand, for every wide span $\mathbf{S}$ of sets, we have the following pair of projections associated with the congruence $G(\mathbf{S})$.

$$G(\mathbf{S}) \underset{\mathrm{prj}_2}{\overset{\mathrm{prj}_1}{\rightrightarrows}} F(S)$$

On the other hand, we have the recombination cones of X (Definition 4.27), which give us a wide span $X(\rho)$ of sets for every cone in ρ in D. It therefore comes quite naturally to our mind that we could coequalize the previous pair of arrows with respect to the wide spans $X(\rho)$ to create a functor $\mathbf{Seg}(\Omega) \to B_2\text{-}\mathbf{Mod}$. Unfortunately, the mapping $\rho \mapsto X(\rho)$ is unlikely to be functorial, so that we need to work a little bit before obtaining such a functor. Specifically, we need to consider the pair of arrows resulting from the following diagram, in which the symbols $\oplus$ denote coproducts in $B_2\text{-}\mathbf{Mod}$ over the finite set of triples (υ, ρ, f) where υ is an object in $\mathbf{Seg}(\Omega)$, ρ is a cone of the form $\Delta_A(\upsilon) \Rightarrow \theta$ in D and f is an arrow $\upsilon \to \tau$ in $\mathbf{Seg}(\Omega)$.

$$\bigoplus_{\upsilon,\rho,f:\upsilon\to\tau} GX(\rho) \underset{\mathrm{prj}_2}{\overset{\mathrm{prj}_1}{\rightrightarrows}} \bigoplus_{\upsilon,\rho,f:\upsilon\to\tau} FX(\upsilon) \xrightarrow{\bigoplus_{\upsilon,\rho,f} FX(f)} FX(\tau) \tag{4.4}$$

Then, for every morphism $g : \tau \to \tau'$, we can define the following diagram, which makes the coequalizer of (4.4) an obvious functor on $\mathbf{Seg}(\Omega)$.

$$\begin{tikzcd}
\bigoplus_{\upsilon,\rho,f:\upsilon\to\tau} GX(\upsilon) \arrow[r, shift left, "\mathrm{prj}_1"] \arrow[r, shift right, "\mathrm{prj}_2"'] \arrow[d, "\subset"'] & \bigoplus_{\upsilon,\rho,f:\upsilon\to\tau} FX(\upsilon) \arrow[r, "\bigoplus_{\upsilon,\rho,f} FX(f)"] \arrow[d, "\subset"'] \arrow[dr, dotted, "\bigoplus_{\upsilon,\rho,f} FX(g\circ f)"] & FX(\tau) \arrow[d, "FX(g)"] \\
\bigoplus_{\upsilon,\rho,f:\upsilon\to\tau'} GX(\upsilon) \arrow[r, shift left, "\mathrm{prj}_1"] \arrow[r, shift right, "\mathrm{prj}_2"'] & \bigoplus_{\upsilon,\rho,f:\upsilon\to\tau'} FX(\upsilon) \arrow[r, "\bigoplus_{\upsilon,\rho,f} FX(f)"'] & FX(\tau')
\end{tikzcd}$$

Definition 4.50 (Recombination semimodules)**.** For every finite chromology (Ω, D) and object τ in $\mathbf{Seg}(\Omega)$, we will denote by $DX(\tau)$ the coequalizer of (4.4). The associated functor $DX : \mathbf{Seg}(\Omega) \to B_2\text{-}\mathbf{Mod}$ will be called the *recombination semimodule over* X

Convention 4.51 (Coequalizer map)**.** For every functor $X : \mathbf{Seg}(\Omega) \to \mathbf{Set}$, the coequalizer map $FX \to DX$ associated with the coequalizer of diagram (4.4) will be denoted as q_X.

Example 4.52 (Recombination semimodule for DNA)**.** Let (E, ε) be our usual pointed set $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}, \varepsilon\}$ and suppose that Ω is the Boolean pre-ordered set $\{0 \leq 1\}$. As usual, we will let $b \in \Omega$ be equal to 1. We here discuss the form of the recombination semimodule defined over the functor $E_b^\varepsilon : \mathbf{Seg}(\Omega) \to \mathbf{Set}$.

If we suppose that the cone $\rho : \Delta_A(\tau) \Rightarrow \theta$ given in Example 4.28 is one of the cones contained in D, then the coequalizer map $q_{E_b^\varepsilon} : FE_b^\varepsilon(\tau) \to DE_b^\varepsilon(\tau)$ will identify the two elements x and y of Example 4.33, which are distinct elements of $FE_b^\varepsilon(\tau)$, as the same element in $DE_b^\varepsilon(\tau)$. In other words, the following identity will hold in $DE_b^\varepsilon(\tau)$.

$$(\mathtt{ACGA}, \mathtt{CG}\varepsilon\mathtt{TT}, \mathtt{TAG}) + (\mathtt{GGTA}, \mathtt{CCTAT}, \varepsilon\mathtt{CA}) = (\mathtt{GGTA}, \mathtt{CG}\varepsilon\mathtt{TT}, \varepsilon\mathtt{CA}) + (\mathtt{ACGA}, \mathtt{CCTAT}, \mathtt{TAG})$$

More generaly, the coequalizer of diagram (4.4) will force any pair (x, y) contained in the recombination congruences resulting from the recombination cones of E_b^ε (see Definition 4.27) to be identified (i.e. $x = y$) in the recombination semimodule $DE_b^\varepsilon(\tau)$.

When the functor $E_b^\varepsilon : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ is a $\mathcal{W}^{\mathrm{bij}}$-pedigrad and ρ is the only cone of the ambient chromology, Theorem 4.47 ensures that the equivalence classes associated with these identifications can be represented by a product operation on the local patches of every element contained in these equivalence classes (see Example 4.48). For chromologies containing more than one cone, the recombination congruences would mix each other so that they would generate recombination relations that could be more refined than those specified by the cones themselves.

4.11. **Logical systems for homologous recombination.** We will denote by $\mathcal{W}^{\mathrm{mon}}$ the class of wide spans $\mathbf{S} = \{S \to S_i\}_{i \in [k]}$ in B_2-**Mod** whose product adjoint arrows $S \to \mathbf{S}^\times$ is a monomorphism in B_2-**Mod**.

Remark 4.53 (Homologous recombination). A $\mathcal{W}^{\mathrm{mon}}$-pedigrad is a pedigrad in which the recombination congruences resulting from the logical system $(B_2\text{-}\mathbf{Mod}, \mathcal{W}^{\mathrm{mon}})$ can be seen as identities (see Proposition 4.70). Another way to put it is to say that a $\mathcal{W}^{\mathrm{mon}}$-pedigrad is a pedigrad in which homologous recombination happens (see Example 4.33).

4.12. **Recombination schemes and pedigrads.** This section determines the right conditions to make a recombination semimodule, as given in Definition 4.50, a $\mathcal{W}^{\mathrm{mon}}$-pedigrad.

Definition 4.54 (Recombination prescheme)**.** We will call a *recombination prescheme*[5] any triple (Ω, D, X) where (Ω, D) is a finite chromology and X is a functor $\mathbf{Seg}(\Omega) \to \mathbf{Set}$.

Example 4.55 (Recombination prescheme)**.** For every finite chromology (Ω, D), element $b \in \Omega$ and pointed set (E, ε), the triple $(\Omega, D, E_b^\varepsilon)$ is an obvious recombination prescheme.

Definition 4.56 (Irreducibility)**.** Let (Ω, D, X) be a recombination prescheme. An object τ in $\mathbf{Seg}(\Omega)$ will be said to be *irreducible* for (Ω, D, X) if for every arrow $f : \upsilon \to \tau$ in $\mathbf{Seg}(\Omega)$, the image $FX(f) : FX(\upsilon) \to FX(\tau)$ coequalizes the following pair of arrows for every cone $\rho : \Delta_A(\upsilon) \Rightarrow \theta$ in D.

$$GX(\rho) \underset{\mathrm{prj}_2}{\overset{\mathrm{prj}_1}{\rightrightarrows}} FX(\upsilon)$$

Remark 4.57 (Coequalizing arrows). Observe that, from the point of view of Definition 4.31, when $\mathbf{S}$ denotes a general wide span, the obvious composite $F(S) \to F(\mathbf{S})^\times \to F(S_i)$ coequalizes the recombination congruence $G(\mathbf{S}) \rightrightarrows F(S)$. This means that if $\mathbf{S}$ is now the wide span given by the ρ-recombination cone of a functor $X : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ (Definition 4.27) for some cone ρ in D, then the morphism

$$FX(\rho_a) : FX(\tau) \to FX(\theta(a))$$

coequalizes the recombination congruence $GX(\rho) \rightrightarrows FX(\tau)$ for every $a \in A$.

[5]The qualifier 'recombination' emphasizes that the fact that the triple (Ω, D, X) is defined with respect to a finite chromology. Any triple of the form (Ω, D, X) where (Ω, D) is not necessarily finite should be called a prescheme.

Example 4.58 (Coequalizing arrows)**.** If we let $\mathbf{S}$ denote the recombination cone $E_b^\varepsilon(\rho)$ of Example 4.28, then Remark 4.57 implies that the recombination congruence $GE_b^\varepsilon(\rho) \rightrightarrows FE_b^\varepsilon(\tau)$ is coequalized by the image of the following arrows via the functor $FE_b^\varepsilon : \mathbf{Seg}(\Omega) \to \mathbf{Set}$.

$$\begin{aligned}
\rho_{a_1} &: (\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) \longrightarrow (\circ\circ\circ)(\circ\circ)(\bullet\bullet\bullet\bullet)(\circ\circ\circ\circ\circ)(\circ\circ\circ)(\circ)\\
\rho_{a_2} &: (\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) \longrightarrow (\circ\circ\circ)(\circ\circ)(\circ\circ\circ\circ)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ)\\
\rho_{a_3} &: (\bullet\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet\bullet)(\circ\circ\circ)(\circ) \longrightarrow (\bullet\bullet\bullet)(\circ\circ)(\circ\circ\circ\circ)(\circ\circ\circ\circ\circ)(\circ\circ\circ)(\circ)
\end{aligned}$$

Indeed, we can check that these arrows would send the pair of equivalent elements

$$x = (\mathtt{ACGA}, \mathtt{CG}\varepsilon\mathtt{TT}, \mathtt{TAG}) + (\mathtt{GGTA}, \mathtt{CCTAT}, \varepsilon\mathtt{CA}) \quad y = (\mathtt{GGTA}, \mathtt{CG}\varepsilon\mathtt{TT}, \varepsilon\mathtt{CA}) + (\mathtt{ACGA}, \mathtt{CCTAT}, \mathtt{TAG})$$

that was given in Example 4.33 to the same elements, which are displayed below.

via $FE_b^\varepsilon(\rho_{a_1})$	via $FE_b^\varepsilon(\rho_{a_2})$	via $FE_b^\varepsilon(\rho_{a_3})$
$\mathtt{ACGA} + \mathtt{GGTA}$	$\mathtt{CG}\varepsilon\mathtt{TT} + \mathtt{CCTAT}$	$\mathtt{TAG} + \varepsilon\mathtt{CA}$

As will be shown in Proposition 4.59, this means that the codomains of the arrows ρ_{a_1}, ρ_{a_2} and ρ_{a_3} may be good candidates for being irreducible objects with respect to a recombination prescheme of the form given in Example 4.55.

Proposition 4.59. *Let (Ω, D, X) be a recombination prescheme. Suppose that D contains a unique cone of the form $\rho : \Delta_A(\tau) \Rightarrow \theta$. For every object $a \in A$, the object $\theta(a)$ in $\mathbf{Seg}(\Omega)$ is irreducible.*

Proof. Let us now show that if ρ is the only cone of D, then the object $\theta(a)$ is irreducible for (Ω, D, X). By definition of a chromology (section 2.8), the arrow $\rho_a : \tau \to \theta(a)$ is an arrow in a category of quasi-homologous segments (see Definition 2.8). By Proposition 2.17, this means that this arrow is the only arrow of type $\tau \to \theta(a)$ in $\mathbf{Seg}(\Omega)$. By Remark 4.57, this means that for every arrow $\tau \to \theta(a)$ in $\mathbf{Seg}(\Omega)$, its image $FX(\tau) \to FX(\theta(a))$ coequalizes the pair $GX(\rho) \rightrightarrows FX(\tau)$. Since ρ is the only cone of D, this means that the image of every arrow $\upsilon \to \theta(a)$ in $\mathbf{Seg}(\Omega)$ via the functor $FX : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ coequalizes the pair $GX(\rho') \rightrightarrows FX(\upsilon)$ for every cone $\rho' : \Delta_{A'}(\upsilon) \Rightarrow \theta'$ in D. □

Proposition 4.60. *Let (Ω, D, X) be a recombination prescheme. For every irreducible object τ in $\mathbf{Seg}(\Omega)$, the coequalizer map $q_X : FX(\tau) \to DX(\tau)$ (Convention 4.51) associated with the recombination semimodule over X (Definition 4.50) is an isomorphism.*

Proof. By Definition 4.56 and universality of a coequalizer. □

Definition 4.61 (Recombination scheme)**.** A *recombination scheme* is a recombination prescheme (Ω, D, X) such that for every cone $\rho : \Delta_A(\tau) \Rightarrow \theta$ in D and object $a \in A$, the object $\theta(a)$ in $\mathbf{Seg}(\Omega)$ is irreducible.

The following theorem shows that the recombination semimodule associated with a recombination scheme is a $\mathcal{W}^{\mathrm{mon}}$-pedigrad.

Theorem 4.62. *Let (Ω, D, X) be a recombination scheme. For every cone $\rho : \Delta_A(\tau) \Rightarrow \theta$ in D, the canonical arrow $\iota : DX(\tau) \to \prod_{a \in A} DX(\theta(a))$ is a monomorphism in B_2-**Mod**.*

Proof. By Definition 4.31, the pullback of two copies of the canonical arrow $\pi_{X(\rho)}$ is the recombination congruence $GX(\rho) \rightrightarrows FX(\tau)$. If we denote by $p_1, p_2 : P \rightrightarrows DX(\tau)$ the pullback of two copies of the arrow ι (see statement), the naturality of the coequalizer map

$q_X : FX \to DX$ gives us an arrow $\lambda : FX(\rho) \to P$ making the following diagram commute.

(4.5)
$$\begin{tikzcd}
 & GX(\rho) \arrow[rr, dashed, "\lambda"] \arrow[dl, "\mathrm{prj}_1"'] \arrow[d, "\mathrm{prj}_2"] & & P \arrow[dl, "p_1"'] \arrow[dd, "p_2"] \\
FX(\tau) \arrow[rr, "q_X"] \arrow[dd, "\pi_{X(\rho)}"'] & FX(\tau) \arrow[rr, "q_X"] \arrow[dl, "\pi_{X(\rho)}"] & DX(\tau) \arrow[dd, "\iota"] & \\
 & & & DX(\tau) \arrow[dl, "\iota"] \\
\prod_{a\in A} FX(\theta(a)) \arrow[rr, "\prod_a q_X", "\cong"'] & & \prod_{a\in A} DX(\theta(a)) &
\end{tikzcd}$$

Let us show that λ is an epimorphism in B_2-**Mod** by showing it is orthogonal with respect to the B_2-semimodule B_2 (see Proposition 4.18). First, because (Ω, D, X) is a recombination scheme, Proposition 4.60 implies that the bottom front arrow of diagram (4.5) is an isomorphism. Second, because $q_X : FX(\tau) \to DX(\tau)$ is a coequalizer map, it is orthogonal with respect to the B_2-semimodule B_2 (Example 4.19). These two facts imply that, for every arrow $x : B_2 \to P$, the composite arrows $p_1 \circ x : B_2 \to DX(\tau)$ and $p_2 \circ x : B_2 \to DX(\tau)$ admits lifts $h_1 : B_2 \to FX(\tau)$ and $h_2 : B_2 \to FX(\tau)$ for which the following diagram commutes.

$$\begin{tikzcd}
B_2 \arrow[r, "h_1"] \arrow[d, "h_2"'] & FX(\tau) \arrow[d, "\pi_{X(\rho)}"] \\
FX(\tau) \arrow[r, "\pi_{X(\rho)}"'] & \prod_{a\in A} FX(\theta(a))
\end{tikzcd}$$

Applying the universality property of $GX(\rho)$ on the previous diagram provides an arrow $h : B_2 \to GX(\rho)$ for which the equation $\lambda \circ h = x$ holds. In other words, the arrow λ is orthogonal to the B_2-semimodule B_2. Now, because the equation $q_X \circ \mathrm{prj}_1 = q_X \circ \mathrm{prj}_2$ holds by definition of q_X and because λ is an epimorphism (Proposition 4.18), the two arrows $p_1, p_2 : P \rightrightarrows DX(\tau)$ must be equal (see diagram (4.5)). Because this pair of arrows is also the pullback of two copies of ι, the arrow ι is a monomorphism (Proposition 4.16). □

4.13. **Presentable functors.** Here, the terminology 'presentable' refers to the idea of being the quotient of a free object of a given type. A presentable functor will usually encompass the type of information that can be observed from a *set* of data with which one wants to analyze in B_2-**Mod**. Throughout this section, we will let $(\Omega, \preceq)$ denote a pre-ordered set.

Definition 4.63 (Presentable functors)**.** A functor $P : \mathbf{Seg}(\Omega) \to B_2\text{-}\mathbf{Mod}$ will be said to be *presentable* over a functor $X : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ if it is equipped with a morphism $q : FX \to P$ that is the coequalizer map of a pair of parallel arrows, as shown below, in $[\mathbf{Seg}(\Omega), B_2\text{-}\mathbf{Mod}]$.

(4.6)
$$Q \underset{p_2}{\overset{p_1}{\rightrightarrows}} FX$$

Example 4.64 (Trivial example)**.** For every pointed set (E, ε) and element $b \in \Omega$, the functor FE^ε_b is obviously presentable for the trivial coequalizer diagram given below.

$$FE^\varepsilon_b \underset{\mathrm{id}}{\overset{\mathrm{id}}{\rightrightarrows}} FE^\varepsilon_b$$

Example 4.65 (Recombination)**.** For every pointed set (E, ε), element $b \in \Omega$ and finite chromology (Ω, D), the recombination semimodule $DE^\varepsilon_b : \mathbf{Seg}(\Omega) \to B_2\text{-}\mathbf{Mod}$ (see Definition 4.50) is the coequalizer for the pair given in (4.6). It is therefore presentable for the coequalizer map $q : FE^\varepsilon_b \to DE^\varepsilon_b$

Definition 4.66 (Presentable morphisms)**.** Let (P_1, q_1) and (P_2, q_2) be two presentable functors over two functors X_1 and X_2, respectively. A *presentable morphism* from (P_1, q_1) to (P_2, q_2) is a commutative square on the following form in $[\mathbf{Seg}(\Omega), B_2\text{-}\mathbf{Mod}]$.

$$\begin{tikzcd} FX_1 \arrow[r, "q_1"] \arrow[d, "g"'] & P_1 \arrow[d, "f"] \\ FX_2 \arrow[r, "q_2"] & P_2 \end{tikzcd}$$

Remark 4.67. Any component-wise composition of presentable morphisms is presentable.

Example 4.68 (Mutations as presentable morphisms)**.** Let (E, ε) be a pointed set, b an element in Ω and (Ω, D) be a finite chromology. One of the most trivial[6] types of presentable morphisms of the form $DE_b^\varepsilon \to DE_b^\varepsilon$ (for the structure given in Example 4.65) can arise from morphisms $FE_b^\varepsilon \to FE_b^\varepsilon$ whose mappings are generated by functions $\nu : E \to F(E)$ that assign each element $x \in E$ to a (finite) sum of elements in $F(E)$. The morphism $FE_b^\varepsilon \to FE_b^\varepsilon$ then sends any word of the form $\mathtt{X}_1 \dots \mathtt{X}_n$ in $E^{\times n}$ to the element of $F(E^{\times n})$ represented by the following finite subset of $E^{\times n}$ (as in Example 4.40).

$$\mathrm{Supp}(\nu(\mathtt{X}_1)) \times \cdots \times \mathrm{Supp}(\nu(\mathtt{X}_n))$$

The following table gives an example when (E, ε) is our usual pointed set $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}, \varepsilon\}$.

$\mathtt{X}$	$\nu(\mathtt{X})$	$\mathtt{X}_1 \dots \mathtt{X}_n$	$\mapsto$	$\mathrm{Supp}(\nu(\mathtt{X}_1)) \times \cdots \times \mathrm{Supp}(\nu(\mathtt{X}_n))$
$\mathtt{A}$	ε	$\mathtt{AC}$	$\mapsto$	$\varepsilon\mathtt{A} + \varepsilon\mathtt{C} + \varepsilon\mathtt{T}$
$\mathtt{C}$	$\mathtt{A} + \mathtt{C} + \mathtt{T}$			
$\mathtt{G}$	$\mathtt{T} + \mathtt{C}$	$\varepsilon\mathtt{GT}$	$\mapsto$	$\varepsilon\mathtt{T}\varepsilon + \varepsilon\mathtt{G}\varepsilon + \varepsilon\mathtt{TT} + \varepsilon\mathtt{GT}$
$\mathtt{T}$	$\varepsilon + \mathtt{T}$			
ε	ε			

As one can imagine, this type of morphism can be used to represent DNA mutations. As noticed in Example 3.34, taking (E, ε) to be the pointed set $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}, \varepsilon\}$ would result in mutations that are too systematic to be realistic. As mentioned thereof, a better parameterization of the alphabet could then be used to make these mutations more realistic. The main difference with Example 3.34 is that B_2-modules now allow us to incorporate some polymorphism in the mutation. The table given below associate examples of mappings (on the left) with certain mutation types (on the right) and shows how the sums can be used to generate polymorphic substitution mutations.

Examples of mappings	$\Rightarrow$	Types of mutations
$\mathtt{A} \mapsto \varepsilon$; $\mathtt{C} \mapsto \varepsilon$; $\mathtt{G} \mapsto \varepsilon$; $\mathtt{T} \mapsto \varepsilon$	$\Rightarrow$	Deletion mutations
$\mathtt{A} \mapsto \mathtt{G} + \mathtt{T}$; $\mathtt{A} \mapsto \mathtt{A} + \mathtt{T}$; $\mathtt{A} \mapsto \mathtt{T}$; etc.	$\Rightarrow$	Substitution mutations

As pointed out in Example 3.34, the fact that the element ε always needs to be mapped to itself (by definition of the concept of presentability over E_b^ε), implies that insertion mutations (if not all) should be studied through spans of presentable morphisms.

Example 4.69 (Transcription as presentable morphisms)**.** Another interpretation of presentable morphisms could be transcription processes (see Example 3.32).

4.14. **Presentable pedigrads and their universal properties.** Let $(\Omega, \preceq)$ be a preordered set equipped with a finite chromology structure (Ω, D).

The proposition given below says that if a presentable functor $\mathbf{Seg}(\Omega) \to B_2\text{-}\mathbf{Mod}$ is a $\mathcal{W}^{\mathrm{mon}}$-pedigrad, then every pair of elements contained in any of its associated recombination congruences will be equated in the images of P. In Example 4.71, we will show how, in general, equations can be used to describe biological phenomena.

[6]More exotic examples are possible.

Proposition 4.70. *Let P denote a $\mathcal{W}^{mon}$-pedigrad that is presentable over a functor $X : \mathbf{Seg}(\Omega) \to \mathbf{Set}$ and ρ be a cone in D. The coequalizer map $FX \to P$ coequalizes the recombination congruence associated with the ρ-recombination cone of X (Definition 4.27).*

Proof. Let ρ be a cone of the form $\Delta_A(\tau) \Rightarrow \theta$. The naturality of coequalizer map $q : FX \to P$ gives us the commutative square displayed below, on the right, while the recombination congruence associated with the ρ-recombination cone of X is given next to it, on the left.

$$\begin{tikzcd}
GX(\rho) \arrow[r, shift left, "\mathrm{prj}_1"] \arrow[r, shift right, "\mathrm{prj}_2"'] & FX(\tau) \arrow[r, "q_\tau"] \arrow[d, "\pi_{X(\rho)}"'] & P(\tau) \arrow[d, "\iota"] \\
& \prod_{a\in A} FX(\theta(a)) \arrow[r, "\prod_a q_{\theta(a)}"'] & \prod_{a\in A} P(\theta(a))
\end{tikzcd}$$

Since, by definition, the arrow $\pi_{X(\rho)}$ coequalizes the pair $(\mathrm{prj}_1, \mathrm{prj}_2)$, so does the composite $\iota \circ q_\tau$. Since P is a $\mathcal{W}^{\mathrm{mon}}$-pedigrad, the arrow ι is a monomorphism, which implies that q_τ coequalizes the $(\mathrm{prj}_1, \mathrm{prj}_2)$ itself (see previous diagram). □

Example 4.71 (More equations)**.** A pedigrad may satisfy various types of equation. For illustration, if we let Ω be the pre-ordered set $\{0 \leq 1\}$, then any pedigrad $P : \mathbf{Seg}(\Omega) \to B_2\text{-}\mathbf{Mod}$ is equipped with a morphism of the following form, where the codomain may contain more equations than the domain (while the equations satisfied in its domain must be sent to similar equations in its codomain).

$$P\Big(\ (\bullet\bullet\bullet)(\bullet\bullet\bullet)(\bullet\bullet\bullet)\ \Big) \longrightarrow P\Big(\ (\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet)\ \Big) \tag{4.7}$$

Interestingly, if we let P be presentable over a functor E_b^ε where (E, ε) is taken to be the $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}, \varepsilon\}$, we can give a biological interpretation to the following equations.

1) **Nullomers:** take the word `GACCGTAAG`, which has representatives in each of the objects displayed in (4.7). Even though these representatives look the same, they may be read differently in the domain and the codomain of (4.7). For instance, the word `GACCGTAAG` could be torsion-free in the domain of (4.7) but could be subject to the following equation in the codomain.

$$(\mathtt{GACCGTAAG}) = 0$$

Such an equation could, for example, mean that the segment `GACCGTAAG` is what one calls a *nullomer* [1], namely a short DNA sequence that cannot appear in the genome of a species. On the other hand, the element `GACCGTAAG` would not be equal to 0 in the domain of (4.7) because the codons that compose it would not be specific to nullomers.

2) **Translation:** take the words `AGTCATGGG` and `AGCCACGAT`, which can be viewed as elements living in the objects of (4.7). These two elements could be distinct in the domain of (4.7), but could be sent to the same element in its codomain.

$$(\mathtt{AGTCATGGG}) = (\mathtt{AGCCACGAT})$$

Such an equation could, for example, mean that the codon translations `Ser-His-Gly` and `Ser-His-Asp` of these two words – which are, noticeably, different – give two tripeptides whose properties are seemingly the same [13].

The proposition given below is more of a categorical result showing that presentable pedigrads are endowed with a universal property. This property says that any logical reasoning done in a recombination semimodule DX can be transferred to any pedigrad that is presentable over the underlying functor X (see Remark 4.73).

Theorem 4.72. *Let (P, p) be a $\mathcal{W}^{mon}$-pedigrad for (Ω, D) that is presentable over a functor $X : \mathbf{Seg}(\Omega) \to \mathbf{Set}$. For every morphism $f : P \to Q$ in $[\mathbf{Seg}(\Omega), B_2\text{-}\mathbf{Mod}]$, there exists a*

unique morphism $f' : DX \to Q$ making the following diagram commute.

$$\begin{tikzcd} FX \arrow[r, "f \circ p"] \arrow[d, "q_X"'] & Q \\ DX \arrow[ur, "f'"'] & \end{tikzcd}$$

Proof. According to Proposition 4.70, the coequalizer map $FX \to P$ makes the following diagram commute for every cone $\rho : \Delta_A(\upsilon) \Rightarrow \theta$ in D.

$$\begin{tikzcd} GX(\rho) \arrow[r, shift left] \arrow[r, shift right] & FX(\upsilon) \arrow[r, "p_\upsilon"] & P(\upsilon) \end{tikzcd}$$

Since, for every morphism $g : \upsilon \to \tau$ in $\mathbf{Seg}(\Omega)$, the diagram given below, on the left, commutes, the corresponding diagram on the right must also commute.

$$\begin{tikzcd} FX(\upsilon) \arrow[r, "p_\upsilon"] \arrow[d, "FX(g)"'] & P(\upsilon) \arrow[d, "P(g)"] \\ FX(\tau) \arrow[r, "p_\tau"'] & P(\tau) \end{tikzcd} \quad \Rightarrow \quad \begin{tikzcd} GX(\rho) \arrow[r, shift left] \arrow[r, shift right] & FX(\upsilon) \arrow[r, "FX(g)"] & FX(\tau) \arrow[r, "p_\tau"] & P(\tau) \end{tikzcd}$$

It then follows that the coequalizer map $p : FX \to P$ coequalizes the pair of arrows of (4.4). By universal property, there must exists a unique morphism $f' : DX \to P$ in $[\mathbf{Seg}(\Omega), B_2\text{-}\mathbf{Mod}]$ making the diagram of the statement commute. □

Remark 4.73 (Presentable morphisms)*.* In Theorem 4.72, if one takes f to be the identity on P, then the statement implies that there exists a unique presentable morphism $(DX, q_X) \to (P, p)$ of the following form.

$$\begin{tikzcd} FX \arrow[r, equal] \arrow[d, "q_X"'] & FX \arrow[d, "p"] \\ DX \arrow[r, "p'"'] & P \end{tikzcd}$$

Then, if one takes f to be a presentable morphism $(P, p) \to (Q, q)$, then Remark 4.67 implies that the composite $f \circ p'$, which must be equal to f' by Theorem 4.72, induces a presentable morphism $(DX, q_X) \to (Q, q)$.

5. Mapping functions and Pedigrads

The goal of this section is to show how one can recover the mapping functions used in gene linkage [26, 16, 9] from pedigrads taking their values in the category of B_2-semimodules. Recall that these mapping functions plot the probability of observing a certain number of cross-overs between pairs of markers on distinct chromosomes as a function of their distances on their respective chromosomes. They usually take the form of (non-normalized) cumulative distribution functions e.g. $y = (1 - e^{-2x})/2$ (see [9, 28]). These functions are important because they give a notion of distance between genes by taking into account the frequency of recombination that separate them [26]. This genetic distance can then be used to study genetic diseases [7, 22].

5.1. Probability theory and B_2-semimodules.

The goal of this section is to reformulate concepts pertaining to Probability Theory [14] in the category of B_2-semimodules in order to reformulate the results of [9] in the context of pedigrads. Our notions will however be weaker than the one usually used in the literature (see [14, Page 8]).

Definition 5.1 (Event spaces)**.** For every set S, we will speak of an *event space* over S to refer to a subset $\mathcal{F} \subseteq \mathcal{P}(S)$ of finite subsets of S that contains the empty set and that is stable under binary unions.

Remark 5.2 (Event spaces as B_2-semimodules). Definition 5.1 is equivalent to requiring $\mathcal{F}$ to a be a sub-object of the free B_2-semimodule $F(S)$ in B_2-**Mod** (see Example 4.2 and Remark 4.8), which is to say a B_2-semimodule $\mathcal{F}$ that is equipped with a monomorphism $\mathcal{F} \hookrightarrow F(S)$ in B_2-**Mod**.

Example 5.3 (Haldane's event space)**.** In [9], Haldane considers an event space in which he can count the number of cross-overs between two DNA segments during homologous recombination. In this example, we show how to recover this event space in the context of pedigrads. To do so, let (E, ε) denote our usual pointed set $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}, \varepsilon\}$ and let $(\Omega, \preceq)$ be the Boolean pre-ordered set $\{0 \leq 1\}$ equipped with a finite chromology structure (Ω, D). The cones of D will be specified later on. We will also let b to be equal to $1 \in \Omega$.

To count the number of cross-overs occurring between DNA segments, as Haldane did, we need to give ourselves a finite set of DNA segments of the same lengths that can be recombined. In terms of pedigrads, this would amount to picking an object τ in $\mathbf{Seg}(\Omega)$ and considering a monomorphism in B_2-**Mod** of the following type.

$$B_2 \hookrightarrow DE_b^\varepsilon(\tau)$$

The fact that this arrow is a monomorphism means that the sum of elements picked out by the element $1 \in B_2$, say $\sum_{k=1}^n s_k$, is not equal to the zero element of $DE_b^\varepsilon(\tau)$. In other words, the recombination of the set of DNA segments $\{s_k \mid k \in [n]\}$ is not degenerate.

Now, the event space considered by Haldane can be seen as the pullback of the arrow $B_2 \hookrightarrow DE_b^\varepsilon(\tau)$ along the coequalizer map $FE_b^\varepsilon(\tau) \to DE_b^\varepsilon(\tau)$.

$$\begin{tikzcd} \mathcal{F} \arrow[r] \arrow[d, hook, "\subseteq"'] \arrow[dr, phantom, "\lrcorner"] & B_2 \arrow[d, hook, "\subseteq"] \\ FE_b^\varepsilon(\tau) \arrow[r] & DE_b^\varepsilon(\tau) \end{tikzcd}$$

The resulting sub-object $\mathcal{F} \hookrightarrow FE_b^\varepsilon(\tau)$ corresponds to the equivalence class of the element $\sum_{k=1}^n s_k$ in $DE_b^\varepsilon(\tau)$. Let us give an example. Suppose that τ is the segment appearing at the top of the following (obvious) exactly 1-distributive cone, call it ρ, and suppose that D only contains ρ.

$$\begin{tikzcd} & (\bullet)(\bullet)(\bullet)(\bullet)(\bullet)(\bullet)(\bullet) \arrow[dl] \arrow[dr] & \\ (\bullet)(\circ)(\circ)(\circ)(\circ)(\circ)(\circ) & \cdots & (\circ)(\circ)(\circ)(\circ)(\circ)(\circ)(\bullet) \end{tikzcd}$$

If the monomorphism $B_2 \hookrightarrow DE_b^\varepsilon(\tau)$ picks out the sum $\mathtt{AGTCCTA} + \mathtt{TCCGAAC}$, then the sub-semimodule $\mathcal{F} \hookrightarrow FE_b^\varepsilon(\tau)$ corresponds to the equivalence class of the recombination congruence associated with the ρ-recombination cone of E_b^ε (Definition 4.27), that is to say all the recombination events between the two words $\mathtt{AGTCCTA}$ and $\mathtt{TCCGAAC}$ (see examples below).

$$\underbrace{\mathtt{AGTCCTA}}_{\text{0 cross-over}} + \underbrace{\mathtt{TCCGAAC}}_{\text{0 cross-over}}$$

$$\underbrace{\mathtt{AGTCCTA}}_{\text{0 cross-over}} + \underbrace{\mathtt{TCTCCTA}}_{\text{1 cross-over}} + \underbrace{\mathtt{TCCGAAC}}_{\text{0 cross-over}}$$

$$\underbrace{\mathtt{TCTCCAC}}_{\text{2 cross-overs}} + \underbrace{\mathtt{AGCGATA}}_{\text{2 cross-overs}} + \underbrace{\mathtt{TCCGAAC}}_{\text{0 cross-over}}$$

etc.

Since $\mathcal{F}$ is a finite subset of $FE_b^\varepsilon(\tau)$, the sum of all its elements corresponds to the representative of the equivalence class of $\mathtt{AGTCCTA} + \mathtt{TCCGAAC}$ (see Example 4.38). This representative event is usually called the *sure event* [14, Page 8]. On the other hand, the intersection of the supports of two events in $\mathcal{F}$ may not exist in $\mathcal{F}$ (see below).

$$\Big\{ \underbrace{\mathtt{AGTCCTA}}_{\text{0 cross-over}}, \underbrace{\mathtt{TCCGAAC}}_{\text{0 cross-over}} \Big\} \cap \Big\{ \underbrace{\mathtt{TCTCCTA}}_{\text{1 cross-over}}, \underbrace{\mathtt{AGCGAAC}}_{\text{1 cross-over}}, \underbrace{\mathtt{AGCGAAC}}_{\text{0 cross-over}} \Big\} = \Big\{ \underbrace{\mathtt{TCCGAAC}}_{\text{0 cross-over}} \Big\} \notin \mathcal{F}$$

The fact that intersections of events do not necessarily belong to $\mathcal{F}$ could translate the idea that natural selection may be able to shape the observable results of homologous recombination (e.g. sperm selection, fetal death, selection of hatchlings, etc.).

However, note that in the case where the arrow $B_2 \hookrightarrow DE_b^{\varepsilon}(\tau)$ picks out a sum of two elements, such as $\texttt{AGTCCTA} + \texttt{TCCGAAC}$, it is always possible to uniquely complete an intersection into an element of $\mathcal{F}$ by considering the complementary recombination operations – this element is the smallest element of $\mathcal{F}$ whose support contains that intersection. For instance, the element given below, on the left, can be seen as the corresponding sum, on the right.

$$\texttt{TCCGAAC} \notin \mathcal{F} \qquad \mapsto \qquad \texttt{AGTCCTA} + \texttt{TCCGAAC} \in \mathcal{F}$$

In the case where the arrow $B_2 \hookrightarrow DE_b^{\varepsilon}(\tau)$ picks out a sum of more than two elements, the completion is not necessarily unique.

Definition 5.4 (Bounded)**.** We will say that an event space $\mathcal{F}$ over a set S is *bounded* if it admits a maximum element for the inclusion of subsets of S. Such a maximum will be denoted as $\cup\mathcal{F}$ and called the *sure event*.

Proposition 5.5. *An event space $\mathcal{F}$ over a set S is bounded, if and only if it is a finite subsets of $\mathcal{P}(S)$.*

Proof. Let $\mathcal{F}$ be a bounded event space over S. Because the element $\cup\mathcal{F}$ belongs to $\mathcal{F}$, it is a finite set. Since $\cup\mathcal{F}$ is also the union of all elements in $\mathcal{F}$, the inclusion $\mathcal{F} \subseteq \mathcal{P}(\cup\mathcal{F})$ must holds, so that $\mathcal{F}$ is finite. Conversely, if $\mathcal{F}$ is finite, then the union of all its element is a maximum element. □

Remark 5.6 (Maximum element)*.* From the point of view of B_2-semimodules, the maximum of a bounded event space is the finite sum of all its elements.

Definition 5.7 (Probability measure)**.** Let $\mathcal{F}$ be a bounded event space over a set S. We will speak of a *probability measure* on $\mathcal{F}$ to refer to a function $\wp : \mathcal{F} \to [0, 1]$ such that

1) both identities $\wp(\emptyset) = 0$ and $\wp(\cup\mathcal{F}) = 1$ hold.
2) if $A, B \in \mathcal{F}$ with $A \cap B = \emptyset$, then the identity $\wp(A \cup B) = \wp(A) + \wp(B)$ holds.

Example 5.8 (Haldane's probability measure)**.** This example carries on Example 5.3 and discusses the definition of a probability measure on the event space defined thereof. First, recall that, in [9], Haldane associates the set of pairs of recombined segment that are stable under complementary recombination operation with a Bernouilli distribution [14, Page 12].

$$\begin{array}{ccccccccccccccc} \texttt{T} & & \texttt{C} & & \texttt{T} & & \texttt{C} & & \texttt{C} & & \texttt{T} & & \texttt{A} & + & \texttt{A} & & \texttt{G} & & \texttt{C} & & \texttt{G} & & \texttt{A} & & \texttt{A} & & \texttt{C} \\ \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\ 0 & & 1 & & 0 & & 0 & & 0 & & 0 & & & & 0 & & 1 & & 0 & & 0 & & 0 & & 0 & & \end{array} \qquad \mapsto \qquad p^1 \cdot (1-p)^5$$

In general, the probability p associated with a recombination event (symbolized, in the previous sum, by 1) is equal to a ratio x/n where n is the number of positions where a recombination can occur on the segment and x is the expected number of recombination event on this segment. In the present case, the probability p should be equal to $x/6$.

For its part, the event space $\mathcal{F} \hookrightarrow FE_b^{\varepsilon}(\tau)$ of Example 5.3 can be associated with a more refined probability measure by associating every DNA segment of a recombination event with one half of the probability that it was given in Haldane's case. A pair of segments that is complementary for a recombination operation will therefore be endowed with the same probability measure as in Haldane's case.

$$\underbrace{\texttt{TCTCCTA}}_{\text{1 cross-over}} + \underbrace{\texttt{AGCGAAC}}_{\text{1 cross-over}} + \underbrace{\texttt{TCCGAAC}}_{\text{0 cross-over}} \qquad \mapsto \quad p(1-p)^5 + \tfrac{1}{2}(1-p)^6$$

$$\underbrace{\texttt{TCTCCTA}}_{\text{1 cross-over}} + \underbrace{\texttt{AGCGAAC}}_{\text{1 cross-over}} + \underbrace{\texttt{AGCGAAC}}_{\text{0 cross-over}} \qquad \mapsto \quad p(1-p)^5 + \tfrac{1}{2}(1-p)^6$$

$$\underbrace{\texttt{TCTCCTA}}_{\text{1 cross-over}} + \underbrace{\texttt{AGCGAAC}}_{\text{1 cross-over}} + \underbrace{\texttt{AGTCCTA}}_{\text{0 cross-over}} + \underbrace{\texttt{TCCGAAC}}_{\text{0 cross-over}} \qquad \mapsto \quad p(1-p)^5 + (1-p)^6$$

In this case, we can verify that if two events are disjoint (i.e. $A \cap B = \emptyset$), then the measure of the union of these is equal to the sum of the measures of each of them (i.e. $\wp(A \cup B) = \wp(A) + \wp(B)$). Of course, other definitions of probability measures are possible.

In [9], Haldane eventually considers long DNA sequences that are subdivided in small intervals so that the probability measure on a sum of complementary segments that are subject to exactly t recombination operations is taken to be equal to the following limit.

$$\frac{n!}{t!(n-t)!}\left(\frac{x}{n}\right)^t\left(1-\frac{x}{n}\right)^{n-t} \underset{n\to\infty}{\longrightarrow} x^t\frac{e^{-x}}{t!}$$

It follows that an observable cross-over event, which is given by the union of all the odd recombination events acting on the sum $\texttt{AGTCCTA} + \texttt{TCCGAAC}$, is equal to $(1-e^{-2x})/2$.

5.2. **Toward more formalism.** We can already see in Examples 5.3 and 5.8 that the computation of mapping functions lack of a certain formalism. This lack of formalism is also mentioned in the literature [17, 26].

1) A first instance in which more formalism seems to be needed lies in the fact that homologous recombination between more than two DNA strands is not well-understood from the point of view of mapping functions [17, Page 1033]. Indeed, while such a recombination event could be noticed from a pair of DNA strands, the actual recombination would only be determined by the knowledge of all the strands resulting from it. This is one of the reasons why the event space of Example 5.3 was not required to be stable under intersection. This also suggests that while homologous recombination between pairs of segments can be modelled via regular probability models, multi-locus recombination between more than two segments might need to be studied from a structural point of view. Pedigrads and chromologies could therefore play a certain role toward this prospect.

2) A second instance that could be mentioned is that Haldane's Poisson model also fails at modelling cross-over interference [26]. Cross-over interference occurs when a recombination event influences the number of recombination happening around it. Cross-over interference is measured via the so-called *coincidence coefficient*, which is the ratio of the recombination rate of two intervals A and B (on the DNA segment) over the simultaneous recombination rate on A and B (see [26]). When this ratio is less than 1, the recombination events are more likely to be clustered on the segment while when it is greater than 1, these events are more likely to be evenly spaced [3].

The desire to incorporate cross-over interference in recombination models [26, 17] could motivate the consideration of topologies as one of the varying parameters. This would be something that the language of chronologies could fit quite well. For instance, instead of considering the cone ρ of Example 5.3, we could now consider recombination cones induced by the following cones of segments.

$$\begin{tikzcd}[column sep=small]
 & (\bullet\bullet)(\bullet\bullet)(\bullet\bullet\bullet) \arrow[dl] \arrow[dr] & \\
(\bullet\bullet)(\circ\circ)(\bullet\bullet\bullet) & & (\circ\circ)(\bullet\bullet)(\circ\circ\circ)
\end{tikzcd}
\qquad
\begin{tikzcd}[column sep=small]
 & (\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet) \arrow[dl] \arrow[dr] & \\
(\circ\circ\circ)(\bullet\bullet\bullet\bullet) & & (\bullet\bullet\bullet)(\circ\circ\circ\circ)
\end{tikzcd}$$

The cone given on the left would model the fact that a recombination event happening on an interval can influence another one that is located farther on the chromosome. This would be due to the fact that a chromosome can pack its DNA, so that the physical distance between two DNA patches is not necessarily related to their distance on the chromosome. On the other hand, the cone given on the right could be used to model a spaced recombination rate. Of course, in order to take into account any possible recombination events of this type, one would need to consider a set of such cones (see below).

$$\begin{tikzcd}[column sep=small]
 & (\bullet\bullet\bullet)(\bullet\bullet\bullet\bullet) \arrow[dl] \arrow[dr] & \\
(\bullet\bullet\bullet)(\circ\circ\circ\circ) & & (\circ\circ\circ)(\bullet\bullet\bullet\bullet)
\end{tikzcd}
\qquad
\begin{tikzcd}[column sep=small]
 & (\bullet\bullet)\,(\bullet\bullet\bullet\bullet)(\bullet) \arrow[dl] \arrow[dr] & \\
(\bullet\bullet)(\circ\circ\circ\circ)(\circ) & \dots & (\circ\circ)\,(\circ\circ\circ\circ)(\bullet)
\end{tikzcd}
\qquad
\begin{tikzcd}[column sep=small]
 & (\bullet)(\bullet\bullet\bullet\bullet)\,(\bullet\bullet) \arrow[dl] \arrow[dr] & \\
(\bullet)(\circ\circ\circ\circ)\,(\circ\circ) & \dots & (\circ)(\circ\circ\circ\circ)(\bullet\bullet)
\end{tikzcd}$$

$$\begin{tikzcd}[column sep=small]
 & (\bullet\bullet\bullet\bullet)(\bullet\bullet\bullet) \arrow[dl] \arrow[dr] & \\
(\bullet\bullet\bullet)(\circ\circ\circ\circ) & & (\circ\circ\circ)(\bullet\bullet\bullet\bullet)
\end{tikzcd}$$

These sets of cones could then be organized in the form of a chromology and the study of the related recombination events would be done with respect to the peak of each of these cones. Making the cones vary would amount to studying a category whose objects are recombination schemes.

6. Conclusion

Pedigrads and chromologies provide a language in which it is possible to talk about many biological phenomena, such as homologous recombination, parsing, mutations, duplications, reversions and CRISPR. Pedigrads in the category of B_2-semimodules were shown to provide a framework in which it is possible to talk about recombination events and genetic linkage.

References

[1] C. Acquisti, G. Poste, D. Curtiss, S. Kumar, (2007), *Nullomers: Really a Matter of Natural Selection?*, Salzberg, Steven, ed. PLoS ONE. 2 (10): e1022.

[2] M. Arenas M, D. Posada, (2010), *The effect of recombination on the reconstruction of ancestral sequences.* Genetics. 184 (4), pp 1133–1139.

[3] K. W. Broman, J. L. Weber, (Jun. 2000) *Characterization of human crossover interference*, American Journal of Human Genetics, vol. 66, no. 6, pp 1911–1926.

[4] R. Brown, T. Porter, (2006), *Category Theory: an abstract setting for analogy and comparison*, In: What is Category Theory? Advanced Studies in Mathematics and Logic, Polimetrica Publisher, Italy, pp 257–274.

[5] C. Ehresmann, (1968), *Esquisses et types de structures algébriques.* Bull. Instit. Polit. Iaşi **XIV**.

[6] Y. S. Chen , A. Torroni, L. Excoffier , A. S. Santachiara-Benerecetti, D. C. Wallace , (1995), *Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups*, American Journal of Human Genetics, 57(1): pp 133-149.

[7] N. G. Copeland, N. A. Jenkins, et al., (Oct 1, 1993), *A genetic linkage map of the mouse: current applications and future prospects.*, Science., vol. 262, no. 5130, pp 57–66.

[8] Jonathan S. Golan, (1999), *Semirings and their Applications*, Springer Netherlands, xii+382 pp.

[9] J. B. S. Haldane, *The combination of linkage values, and the calculation of distance between the loci of linked factors*, J Genet, Vol. 8 (1919), pp 299–309.

[10] T. Kivisild, (2015), *Maternal ancestry and population history from whole mitochondrial genomes*, Investigative Genetics, 6:3.

[11] W. Lawvere, *Functorial Semantics of Algebraic Theories.* Ph.D. Thesis, Columbia University, New York, NY, USA, 1963.

[12] Y. Lazebnik, *Can a biologist fix a radio?–Or, what I learned while studying apoptosis.*, Cancer Cell. 2002 Sep; 2(3):179–82.

[13] Yunsheng Li et al., *Dipeptide seryl-histidine and related oligopeptides cleave DNA, protein, and a carboxyl ester*, In Bioorganic & Medicinal Chemistry, Volume 8, Issue 12, 2000, pp 2675-2680.

[14] M. Loève, *Probability Theory*, Vol 1. New York: D. Van Nostrand Company, 1955.

[15] S. Mac Lane, (1998), *Categories for the working mathematician*, Graduate texts in mathematics, Springer, New York.

[16] M. S. McPeek, (1996), *An Introduction to Recombination and Linkage Analysis.* In: Speed T., Waterman M.S. (eds) Genetic Mapping and DNA Sequencing. The IMA Volumes in Mathematics and its Applications, vol 81. Springer, New York, NY.

[17] M. S. McPeek , T. P. Speed, (1995) *Modeling interference in genetic recombination*, Genetics, February 1, vol. 139, no. 2, pp 1031–1044.

[18] D. A. Morrison, February 2009, *Why would phylogeneticists ignore computerized sequence alignment?*, Syst Biol.;58(1):150-8.

[19] D. A. Morrison, October 2009, *A framework for phylogenetic sequence alignment*, Plant Systematics and Evolution, Volume 282, Issue 3–4, pp 127–149.

[20] D.W. Mount, (2004), *Bioinformatics: Sequence and Genome Analysis*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 692 pp.

[21] P. B. Needleman, C. D. Wunsch, *A general method applicable to the search for similarities in the amino acid sequence of two proteins*, Journal of Molecular Biology, Volume 48, Issue 3, 1970, pp 443–453

[22] J Ott, (1 Jan. 1999), *Methods of analysis and resources available for genetic trait mapping*, Journal of Heredity, Volume 90, Issue 1, pp 68–70.

[23] E. Pennisi, (2013), *The CRISPR Craze*, Science, vol. 341, no. 6148, pp 833–836.

[24] A. B. Reams, R. R. John, (3 Aug. 2017), *Mechanisms of Gene Duplication and Amplification.* Cold Spring Harbor Perspectives in Biology 7.2 (2015): a016592. PMC. Web.

[25] M. Rosenberg (Ed.), (2009), *Sequence Alignment: Methods, Models, Concepts, and Strategies.* University of California Press.

[26] T. P. Speed, (2005), *Genetic Map Functions*, Encyclopedia of Biostatistics, 2nd Edition, John Wiley & Sons, Ltd.
[27] D. I. Spivak, *Category theory for the sciences*. MIT Press, Cambridge, MA, 2014. viii+486 pp.
[28] H. Zhao and T. P. Speed, (Apr. 1996), *On Genetic Map Functions*, Genetics, vol. 142, no. 4, pp 1369–1377.
[29] A. Ziegler, I. R. Knig, (2006), *Genetic Distance and Mapping Functions*, John Wiley & Sons, Ltd.
[30] P. H. Sellers, *Pattern recognition in genetic sequences by mismatch density*, Bulletin of Mathematical Biology, July 1984, Volume 46, Issue 4, pp 501–514.

M.I.T., Department of Mathematics, 77 Massachusetts avenue, Cambridge, MA 02139
E-mail address: rtuyeras@mit.edu