DNA Sequences from Below: A Nominalist Approach

Yu Lin[1], Peter Simons[2]

1. Genome Resource and Analysis Unit, Genomics Laboratory, Center for Development Biology, RIKEN, Japan linyu@cdb.riken.jp
2. Department of Philosophy, Trinity College Dublin, Ireland psimons@tcd.ie

Abstract. We define *DNA sequence* by a bottom-up approach, starting with a real sequence from an actual biological sample. By providing axioms for notions of string, substring and strand, we formally define a DNA sequence, and a DNA molecule as composed of two anti-parallel strands. We note that a sequence is a kind of group in which each member stands a certain relation to every other. The spatial aspects of a DNA sequence are also described.

Keywords: DNA sequence, string, group, spatiality, nominalism

DNA sequence in molecular biology

Molecular biology is the study of biology at the molecular level. It has roots in biochemistry, genetics, microbiology and crystallography for molecular structures. The discipline is determined by two historic scientific events: the discovery of the double-helix structure of DNA in 1953, and the completion of the sequence of the human genome in 2003. The double-helix structure has implications far beyond the structure itself. It showed that "the secret of the gene would be a linear, digital message", as Matt Ridley summarized in his forward to *Inspiring Science Jim Watson and The Age of DNA*.[1] As a result, modern biologists may manipulate a whole genome sequence using a computer. Molecular biology is growing and converging with other disciplines: medicine, engineering, computer science, and the humanities. Its central task however has never changed since its beginning: to elucidate the interrelations of DNA, RNA and proteins.

In Wikipedia (http://en.wikipedia.org/wiki/Genetic_sequence), it is stated that "A DNA sequence or genetic sequence is a succession of letters representing the primary structure of a real or hypothetical DNA molecule or strand, with the capacity to carry information as described by the central dogma of molecular biology." We will be seeking to improve on this definition, which for a start contains what

philosophers call a use–mention error: it confuses the sequence itself with the letters conventionally used to represent the bases.

1.1 Sequencing process

DNA sequences can be derived from the biological raw material through a process called DNA sequencing. Since the formal initiation of the Human Genome Project in 1990, DNA sequencing has entered a new era powered by supercomputers. So-called "shotgun sequencing", proceeds by first randomly cutting the target sequence (up to a billion base pairs) into smaller fragments. After putting the fragments into a sequencer that will generate raw readings for fragments, a computer then attempts to combine all the readings obtained from the sequencer to give the whole sequence, which is called an assembly process.

There are different technologies for the sequencing process. Currently, advanced state-of-the-art next-generation sequencing platforms such as the Roche-454 GS FLX, Illumina Genome Analyzer and ABI SOLiD provide high-throughput and high-speed technology in reading the nucleotide bases of samples. However, the reading length generated by such sequencers is very short: ~400bp by Roche-454 FLX titanium, ~75bp by Illumina, and ~50bp by ABI SOLiD. A more old-fashioned but stable and widely used sequencer is ABI 3130/3730 series, by which a reading length up to 2000bp can be obtained.

Both raw readings generated by sequencers and assembled sequences are composed of four letters: A, C, G, and T, representing the four nucleotide bases of a DNA strand — adenine, cytosine, guanine, and thymine — covalently linked to a phosphodiester backbone. In the typical case, the sequences are printed abutting one another without gaps, as in the sequence 5'AAAGTCTGAC3', with a reading direction from 5' to 3'. Sometimes, due to technological limitations, a next-generation sequencer cannot detect the type of nucleotide in a certain position, such as a repetitive sequence called polyA: 5'AAAAAAAAAAAAAAAAAAA3'. Then, N is used for representing any of the four nucleotide bases; in the case of polyA, 5'AAAAANNNNNNNNNNNNN3' could be the raw reading generated by a next-generation sequencer.

The completion of the sequence of a genome of an organism means that the completion of the full picture of the genome constructed from the puzzle of millions pieces of sequences generated by a modern sequencing method. There are two concrete conditions that need to be satisfied:

1. all the pieces need to be located in the correct chromosomes;
2. all the pieces need to be put in the correct order within each single

chromosome.

DNA sequences are concrete

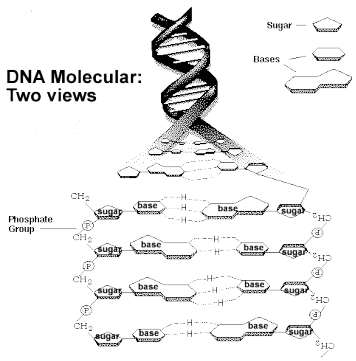
There are at least two kinds of nominalism, one that maintains that there are no universals and one that maintains that there are no abstract objects. [2] The nominalism we employ here maintains that no *abstract objects* need be considered in regard to DNA sequences in molecular biology. This will be shown possible by actually doing it.

1.2 DNA sequences are not abstract objects

A standard conception of abstract entities is that they are non-spatial (or non-spatiotemporal) and causally inert items. [3]
A DNA material isolated from a real biological sample, for example, a drop of blood from a bird, is in a straightforward sense a spatial temporal concrete thing. The DNA sequence of this material is its primary structure, a linear ordered set linked by its nucleotide bases: A, T, C, G, such as 5'AATTCTGATAAGAA3'. However, after the sequencing process, this material will be broken down. The raw readings generated by sequencers on this sample are not abstract either, because they reside in an electronic file with a physical location in a computer's file system. The same assertion could be made with the file of the complete sequence which has been assembled by software from small fragments. The inner connection of those three is such that the raw readings or the assembled sequence reflect the base order of the DNA material.

1.3 Direction of DNA sequence is essential

As shown in fig1, the DNA molecule's primary double-helix structure's essential element is "the manner in which the two chains are held together by hydrogen bonds between the bases", as J. D. Watson and F. H. C. Crick pointed out in 1953. The two chains of the DNA molecule run in opposite directions. The necessity for specific base pairing, e.g. adenine must be paired with thymine in the opposite chain, "demands a definite relationship" between the sequences of the bases on the two chains. Therefore, each chain will be the



complement of the other, and "if we know the actual order of the bases on one chain, we could automatically write down the order of the other". [4]
There is a physical and chemical sense in which direction matters. The convention of writing sequences of bases from the 5' end to the 3' end of a sequence has a good scientific basis. It is also the only direction in which synthesis *in vivo* can take place.

1.4 One sequence, three different things

Although a DNA molecula is in a double-helix structure, because of its specific base paring, one chain's base order is enough to give all the information about the DNA molecula. However, depending on different contexts, a given DNA sequence can have analogous meanings.
For example, given a DNA sequence (5')TTGCAGTATTTAATT(3'), we can use it to represent three different things:

1) The order of the bases in a sequence forming of a part of a real DNA molecule. It represents the primary structure of the double-helix in a simplified way. The direction of 5'->3' is necessary.
2) The letters of a raw reading generated by a sequencer. Sometimes, if generated by a next-generation sequencer, directional information may be null.
3) Assembled sequences, sometimes contigs, generated by software that calculate overlaps of raw readings. Sometimes there will be no direction for this sequence.

As we claimed before, all these three types are concrete things with a spatiotemporal aspect.
The first is a kind having a different property from the other two. It represents a *real sequence* as a biological material entity, which has a directional physical location on a specific chromosome. There is no gap in the sequence, which means there will be no other letter than A, T, C,or G to represent an element of the sequence.
The other two kinds 2) and 3) are information entities which reflect part of the reality of the first one. We might call these the "sequences" of a sequence. (Here, "sequence" means a text string rather than a real sequence.) In these kinds, N or other letters [5] can be used for a gap or an undetectable nucleotide in a position.

Nominalism about the real sequence

To give a formal account of the real sequence, we start from the ideas of element and sequence rather than element-kind and sequence-kind.

We assume there are some individuals which (because they are basic for genomic purposes) we call *elements*, and that each element belongs to exactly one element-kind, so that no two element-kinds have any element in common. Two element-kinds are called *coextensional* when the same elements belong to them. Element-kinds with at least one element in common are therefore coextensional.
Some elements occur in what we may call sequences. A sequence consists literally of several elements linked together one after the other. How they are linked is not part of the abstract theory except that the linking is assumed to be real (physical) and asymmetric, i.e. if one element **A** is linked to another **B**, the short sequence **AB** is such that there can be no sequence **BA** of the elements in the opposite order (at the same time). How things turn out later if the elements of a sequence are rearranged is a different matter.
We assume without further justification that the physical links between elements can all be treated as of the same kind. This may be a simplification in some cases but for the purposes of defining sequence we overlook it. In the case of DNA strands the linking is provided by bonding between carbon atoms in the phosphate deoxyribose backbone. The directionality is given by the difference between the 5' and 3' positions of carbon in a deoxyribose ring, giving rise to the upstream/ downstream asymmetry in DNA.

A finite whole W of elements E

We define a whole of several elements **E** as an object which

1. has every element of **E** as its part

2. has no other types of elements than **E** as its parts

It may or may not have non-element parts such as parts of the backbone which are not bases in DNA. For a definition of sequence we do not need to take these into account.

A finite whole **W** of elements **E** is one where there are finitely many elements in **E**. In all practical applications this will be the case. Infinite collections may interest mathematicians but not real scientists.

Let **W** be a whole of elements.

Let **El(W)** be the collection of elements,

W is a single individual of which all elements of **El(W)** are parts.

If an individual **A** is one of the elements of **W** we write:

A is one of **El(W)**.

The relation between **A** and **W** thereby defined is the element-of relation:

A is an element of **W.**

Note that if **A** is an element of **W** implies that **A** is a part of **W**. But not necessarily *vice versa*: **W** typically has other parts than its elements.

Definition of L-string

We say **W** is a *string* of its elements **E** iff (Def.) there is a relation defined on the elements **E** which is connected, asymmetric, bi-directionally unique or non-splitting, and has two exactly free ends.

1.4.1 Linking relation L

Let **A** and **B** both be elements of **W**:

If **A** links to **B**, we write '**ALB**' .

We are using the schematic letter '**L**' for such a linking relation. What the link consists of in any concrete case may vary. In the case of a DNA sequence it is the usual phosphodiester bonds between the fifth and third carbon atoms of adjacent sugar rings in the molecule. In the conventional notation for DNA the "upstream" base is written to the left of the "downstream" base and we are reflecting this in our choice of notation. We also stipulate that if **ALB** we may say that **A** is (immediately) to the left of **B** and **B** is (immediately) to the right of **A**. The use of 'left' and 'right' is here merely conventional and corresponds to no actual left–right orientation (which is in any case observer- and orientation-relative).

L is assumed to be asymmetrical. Ignoring the temporal aspect, the axiom of Asymmetry says:

Asym. For all **x** and **y** in **E**: if **xLy** then not **yLx** . (**0**)

Corollary:

For no **x** in **E** is **xLx**.

Above axioms ensure the local direction of a sequence is from **x** to **y**.

If **ALB** then there exists a sequence of **A** linked immediately to **B** which we write '**AB**'. **A** and **B** are both parts of **AB** and neither is

identical to **AB**: they are proper parts. The asymmetry tells us that

If **AB** exists **BA** does not exist .

We may extend the idea of linking from single elements to longer sequences of elements **E,** the idea being that one sequence **S** is **L**-linked to another **T** when the rightmost element of **S** is **L**-linked to the leftmost element of T. This explains what happens when we "paste"' together short subsequences resulting from modern sequencing methods.

1.4.2 Non-splitting or branching

Non-branching sequence from the left:

LUn. For all **x**, **y** and **z** in **E**: If **xLy** and **zLy** then **x** = **z** . (2)

Non-branching sequence from the right:

RUn. For all **x**, **y** and **z** in **E**: If **xLy** and **xLz** then **y** = **z** . (3)

1.4.3 Two free ends

A sequence is left-ending:

LEnd. For some **x** in **E** there is no **y** in **E** such that **yLx** . (4)

A sequence is right-ending:

REnd. For some **x** in **E** there is no **y** in **E** such that **xLy** . (5)

1.4.4 Ancestral of L-relation

To exclude other cases such as two unconnected strings, or one string and one cycle in **W**, we require that the string be *connected,* that is, that it be possible to pass from any element to any other element by a finite sequence of steps in the **L**-direction or in the direction opposite to **L**.

For this purpose, we start from a definition of the *ancestral* of the **L**-relation.

Intuitively, **A** stands in the ancestral of the **L**-relation to **B** if **ALB** or **ALx** and **xLB** for some **x** or **ALx** and **xLy** and **yLB** for some **x** and **y** … and so on. Thus, we get from **A** to **B** by doing one **L**-step after another.

To define ancestral we start by taking **C** to be a collection of elements.

C is **L**-hereditary: Her(**L**)(**C**) iff (Def.) for all **x** and y: if x is one of C and **xLy** then y is one of **C**.

So **C** collects up all the things lying **L**-downstream of **x**. We now define the ancestral **L*** of **L** as follows:

For all **x** and **y**: **xL*y** iff (Def.) for all **C**, if **x** is one of **C** and Her(**L**)(**C**) then **y** is one of **C**.

This takes all the **L**-hereditary collections and extracts the smallest or minimal one, since an arbitrary **L**-hereditary collection may contain extraneous junk. If **xL*y** then **y** is **L**-downstream of **x** by a finite number of steps.

> **L**-Conn. A collection **C** of elements is **L**-connected iff (Def.)
> for all **x** and **y** in **C**, either (i) **x** = **y** or (ii) **xL*y** or (iii) **yL*x**. (**5**)

A whole made of an **L**-connected collection of elements may not contain two disconnected sub-collections, but it could be cyclic.

Therefore, we give a final definition for an **L**-string object:

An **L**-*string* of elements **E** is a finite whole made of elements linked by the **L**-relation such that it satisfies Asym, LUn, RUn, LEnd, REnd and L-Conn.

Substring and L-strand

One string is a substring of another under the following conditions.
Let **S** be a string of elements **E** under linking relation **L** and **T** be a string of elements **F** under linking relation **M**.

We say **S** is a substring of **T** iff (Def.)
1. **S** is a string of **E** under **L**
2. **T** is a string of **F** under **M**
3. **L** = **M**
4. **E** is a sub-collection of **F**, i.e. every one of **E** is one of **F**.

We now define an **L-strand** of elements **E** as a maximal L-string.

Let S be an L-string of elements **E.** Then **S** is an **L-strand** iff (Def.) there is no **T** and no **F** such that **T** is an L-string of elements **F** and **F** contains all elements of **E** and **S** is a substring of T.

We then define an **L**-strand having two ends:
Let **S** be an **L**-string, then **L**-top and **L**-bottom elements of **S** are the

two ends respectively:

LTop(**S**) = that element **x** of **S** such that for no **y** in S is **yLx**

LBot(**S**) = that element **y** of **S** such that for no **x** in **S** is **yLx**

Sequences or string-kinds

We assume every element **A** belongs to at least one element-kind **EK(A)**, and that if two element kinds **K** and **H** have any element in common, they are one and the same element-kind. Therefore, distinct element-kinds are disjoint, i.e. have no common members (for if they did, some element would belong to two distinct kinds). Thus, every element belongs to exactly one element-kind.

Let **A** and **B** be elements. **A** and **B** are *isogenic* iff they belong to the same element-kind.

Let **S** and **T** be two **L**-strings. It is supposed that they are equally long, i.e. the collection **El(S)** and **El(T)** have the same number of members. We then define a correspondence relation **L**-*corr* between elements of the strings as follows:

(i) **L**Top(**S**) **L**-corresponds to **L**Top(**T**) and vice versa, and neither **L**-corresponds to anything else

(ii) For all **x** and **y** in S and for all **z** and **w** in **T**:
if **x** **L**-corresponds to **z** and **xLy** and **zLw** then **y** **L**-corresponds to w and vice versa and neither **L**-corresponds to anything else.

By this means we set up a one-one correspondence between elements of **S** and the elements of **T** that correspond to them in position down from the top. First in **S** corresponds to first in **T**, second in **S** to second in **T** and so on.

We now define *isotypy* between two **L**-strings.

Let **S** and **T** be two **L**-strings. Then **S** is **L**-*isotypic* with **T** iff (Def.)
(i) **S** and **T** are equally long.
(ii) For all **x** in **S** and **y** in **T**, **x** is isogenic to **y** (they are of the same element-kind), and **x** **L**-corresponds to **y**

We may now say: two **L**-strings **S** and **T** exemplify the same string-type if and only if they are **L**-isotypic.

We may specify a string-type by giving:
(1) The linking relation **L**
(2) The number of elements in any instance

(3) The element-kind of the **L**-top of each instance
(4) The element-kind of each subsequent element as we pass along the string, or equivalently
(4') For each position in the string-kind, the element-kind of the occupier of that position in any instance.

The positions in a string-kind derive from the number of **L**-steps that we go from the **L**-top of any string of that kind to the element that number of steps along. First, second, third and so on. These can be notated by the standard numerals.

Pairs in a DNA molecule

The above notion of correspondence is not the same as that of pairing used in genetics. There we talk about the actually linking of distinct elements in the two strands of DNA that form up to make a double helix, such that G pairs with C and A with T. What this means is that not just any conceptually possible strand of A, C, G and T bases can be a DNA strand in a double helix DNA molecule, because the two strands have to be anti-parallel. That imposes a constraint on the sequence of bases, as follows:
Call paired element type *counterparts*: so A and T are each other's counterparts, and G and C are each other's counterparts.

A strand of bases is only a candidate for a double DNA strand if:

for all $n \geq 0$: the position n **L**-steps *down* from **L**Top (first position, 5' end) of one strand is in its element-kind the counterpart of the position n steps *up* from **L**Bot (last position, 3' end) of the other strand.

We call such a sequence *Definitely Nicely Arranged*.

Two Definitely Nicely Arranged and isotypic strands can pair in an anti-parallel fashion to form a complete DNA molecule: any actual DNA molecule has two strands which are DNA (Definitely Nicely Arranged), isotypic, and do actually pair up by hydrogen bonds.

Discussion

1.5 Group

A sequence is one kind of group, in the sense of some kind of collective entity. We can also call these 'collections', but the terminology is not so important. There are lots of kinds of groups in this sense. Examples of

such groups include orchestras, football teams, herds of cattle, galaxies and so on.
As an example consider an archipelago, which is a geographical group of islands. One preferred definition of an archipelago is:

An archipelago is a group of several islands in a single expanse of sea such that each island is closer to at least one other of the same group than it is to any island not of that group.

The point is that what makes something an archipelago is that

(1) It has several members, each of which is an island.
(2) Each of these islands stands in certain relations to the other members in which they do not stand to other islands.

So there are three requirements:

(1) There are several members (more than one)
(2) Each of the members is of a certain kind (here: island)
(3) There are certain relations among these members

If the several objects *m* are the members, *k* is their kind, and *r* are the relations in which they stand, the group is fully determined by these three aspects. Now we are used as logicians to treating a kind *k* as a sort of relation (one with only one place), so let's subsume the kind and any other required properties among the relations *r*. Then we say that the group *m;r* is specified entirely by *m* and *r*. What that means is that:

A group *m*;r** is the same group as *m;r* iff $m = m^*$ and $r = r^*$.

Therefore we say a sequence is a group of which every member stands the *L-string* relation with each other.

1.6 DNA sequence as formal

One has to be careful in saying that the notion of sequence is formal. That is true, but sequences of certain kinds, that is, groups which are sequences but also something else, are not generally of their kind formally.
Taking a DNA sequence, the members have to be base molecules of the four kinds, and the links are the standard chemical ones. There is directionality to the links between successive bases. The DNA sequence has to have a first and a last member, no gaps, no loops, no splitting. It is linear with ends. Thus, what makes a DNA sequence a *sequence* is formal. What makes it a *DNA* sequence is material, i.e. special and particular.

A DNA sequence can be modeled by letters or numbers or many other things, and provided the model retains the number and order and distinction of kinds represented, it is a true one. This is because the sequence is formal: it's like a long "word" made of an alphabet of four letters. Such a "word" has exactly the same *formal* structure as the DNA molecule, but is of a different *material* kind, consisting of letters adjacent on the page without bases linking them chemically. (By 'material' here we don't mean 'made of matter' but simply 'not formal'.) The difference between letter sequences and DNA sequences is thus material, not formal.

1.7 Spatial aspect of a DNA sequence

There are many temporal and spatial sequences existing in the world. In a DNA sequence the relation is partly spatial because successive bases are spatially next to one another, as well as chemically bonded (two bases that are not close together cannot be chemically bonded).
Often temporal and spatial sequences are interlinked in important ways. For example in DNA the upstream to downstream order of bases, which is a spatial-chemical sequence, is reflected in the temporal order of the *in vivo* synthesis process, in that first the upstream base is added, then the next downstream base and so on. In this case the temporal order is not that of the base but that of an event of a new base molecule attaching to the partial strand.

1.8 Sequence Ontology: another formalization of sequence

Sequence Ontology (SO) [5] was initiated in 2003 by the Sequence Ontology Consortium, a joint effort by genome annotation centers, including WormBase, FlyBase, the Mouse Genome Informatics group and the Sanger Institute. The purpose of Sequence Ontology is similar to ours, that is, to formalize the notion of genomic sequence. It has been applied to integrating genomic data from different databases, as well as semantic interoperability for both human and machine.
Hoehndorf R. et al. developed an axiom system in predicate logic that serves as a foundation for the top-level categories of SO: *Sequence* and *Junction*. [6] Most of the axioms are available in first-order logic, however, some definitions, such as, connectivity, are in second-order logic. In this system, there is another class *Molecule* that has been defined as a superclass of sequence tokens. Also, SO claims that multiple molecules can have the same sequence, and a sequence exists as long as there is a molecule with that sequence.

This theory is different from ours, because we took a sequence as a real part of a DNA molecule that exists during a certain time period and occupies a concrete space. Another feature in our theory is that the electronic documents of a sequence, i.e., a record of an EST sequence with a NCBI identifier, is another type of sequence similar to the words written in a book. As nominalists, we deny that the abstract sequence classification reflects the ground-level facts.

Currently, in the SO ontology developing community, there is confusion over the definitions of some classes, such as whether a class will be classified as a molecule or a sequence. [7] For example, in SO, polypeptide and transcript are described by genomic context, that is the region of the genome that encodes their sequence. The 'super-classes' hierarchy of polypeptide is as following:

- sequence_feature
  -region
    -biological_region
      -polypeptide

The definition of polypeptide is: "A sequence of amino acids linked by peptide bonds which may lack appreciable tertiary structure and may not be liable to irreversible denaturation".

The definition of polypeptide given by SO is a biological one rather than a formalized logical one. According to our theory, a polypeptide is a kind of sequence whose elements are amino acids satisfying the L-String axioms.

SO uses a subsumption hierarchy to describe the kinds of features and a mereonomy to describe their part-whole structures. Sequence features are related by their genomic position. Thus, a *sequence _feature* is an extended or non-extended biological sequence. Extended sequence features are regions such as genes, intergenic regions or sequences of polypeptides, whereas non-extended sequences are junctions: the boundary between two extended sequences. The definitions of terms are currently being refined in SO. [7]

Here, we started by giving the definition of the most primitive term in molecular biology, a DNA sequence. We gave the definition for a sequence as well, thus, the class *region* in SO is a subclass of sequence in our theory. We believe that our definition will elucidate some ambiguous classifications and inconsistency in the current SO community.

Of course, there are many other more definitions in SO which we did not discuss in this paper, e.g., the class *Junction* in SO, which is actually a *Boundary* type that has been formally defined in BFO.[8] More new terms will appear as scientific investigations in molecular biology continue. In attempting to achieve a formal theory of molecular

biological genomic entities, such as variation, transcript, gene and so on, more mappings and collaboration with the SO community need to be carried through in the future.

Conclusion

In this paper, we claim that a DNA sequence of a DNA molecule is a spatial-chemical sequence of the nucleotide bases. It is a formal spatial sequence. A sequence is a kind of group in the sense of a collective, and each member stands a certain relation with each other. In the case of DNA sequence, the L-string relation we described is the relation which makes members (elements) of a DNA sequence, individual instances of the four base kinds, into a formal sequence. In a nominalistic framework, we first defined the **L**-string axioms, then substring and sequences by a bottom-up approach. At no stage do we employ abstract objects such as sets, numbers of functions in our definitions, but only the ideas of concrete individuals, collections (groups) of concrete individuals, and concrete relationships among individuals, such as those actually obtaining between adjacent bases in real DNA molecules.

We must warn against using the term 'sequence' when it is not appropriate. Sometimes things are ordered but not in a sequence. For example, the points on a straight line are ordered, but not in sequence. The reason is that in a sequence there is a first, a second, and so on, and for each term except the last term there is a next term, like in the numbers 1, 2, 3, ... If there are uncountably infinitely many members in a group, such as the points on a line, we should not use 'sequence' in this case.

Our theory focusses on formalizing molecular biological sequences in the genomics context. It represents initial work and could be extended to both domain-neutral and domain-specific levels. Application for this theory has been tested for assembling contigs to a completion of a 67kpb's target sequence in actual sequencing practice, which was introduced in detail in [9], in which an Ontology for Genetic Interval (OGI) [10] and its relations have been employed.

Note YL: This paper is inspired by discussion with Peter Simons when we met in InterOntology09, February 2009 in Tokyo. Most of the content comes from PS's emails to me about DNA sequence, groups, temporal or spatial aspects of sequence, and so on. I am responsible for asking questions on above issues. P.S. is the main contributor for the content of logic axioms and metaphysics discussion in this paper.

References

1. Inglis J.R., Sambrook J., Witkowski J.A.: Inspiring Science: Jim Watson and The Age of DNA. Cold Spring Harbor Laboratory Press. xv-xix (2003)
2. Rodriguez-Pereyra, G.: Nominalism in Metaphysics. The Stanford Encyclopedia of Philosophy *(Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/nominalism-metaphysics/>
3. Rosen, G.: Abstract Objects. The Stanford Encyclopedia of Philosophy (Fall 2009 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2009/entries/abstract-objects/>.
4. Watson, J.D., and Crick, F.H.: The structure of DNA. Cold Spring Harbor. Symp. Quant. Biol. 18: 123–131 (1953)
5. www.sequenceontology.org/
6. Hoehndorf R., Kelso J., Herre H.: A Formal Ontology of Sequences. Available from Nature Precedings <http://dx.doi.org/10.1038/npre.2009.3537.1> (2009)
7. Eilbeck K. and Mungall C.: Evolution of the Sequence Ontology terms and relationships. Available from Nature Precedings <http://dx.doi.org/10.1038/npre.2009.3495.1> (2009)
8. Basic Formal Ontology http://www.ifomis.org/bfo/
9. Lin Y., Tarui H., Simons P.: From Ontology for Genetic Interval(OGI) to Sequence Assembly – Ontology apply to next generation sequencing. Proc. SWAT4LS Workshop, Amsterdam, Nov.20th, 2009. (in press)
10. Ontology for Genetic Interval http://bioportal.bioontology.org/ontologies/40117