Procedurally Generated Artworks Based on Multiple Sequence Alignment of Orthologous Gene Copies

MARTIN CALVININO

In this article, the author presents a novel approach to the procedural generation of artwork series based on multiple sequence alignment of orthologous gene copies. In the strategy developed, nucleotides present in a string of DNA (A, G, C, T) were each assigned to an existing artwork. New visual compositions were then created by collaging columns of pixels from each of the existing four artworks according to the arrangement of nucleotides after orthologous genes were aligned. The resulting outcome was a distinctive set of artworks in which visual differences were governed by nucleotide divergence at genes of interest due to evolutionary processes.

GENOMICS CONVERGES WITH COMPUTATIONAL ART

The sequencing of plant genomes from related grasses that share a common ancestor allows for the investigation of sequence divergence at chromosomal segments that have retained their gene content and order throughout speciation [1–3]. The field of science that studies the evolution of genomes from related plant species is known as plant comparative genomics and has been an area of study for me during my years as a practicing scientist within academic settings [4–8].

When plant comparative genomics is approached from the perspective of computational art—my current field of study—the processes and methods traditionally used in scientific research can be co-opted for art creation as means for visual experimentation and self-expression [9,10]. I have focused my efforts on the integration of plant genomics with computational art, producing artworks inspired by plant genome browsers [11], rice chromosomal segments [12], maize genome evolution [13], the auditor perception of nucleotide sequence contexts and reduction in genome divergence as consequence of plant domestication [14,15], and the creation of synthetic microRNA genes using machine learning algorithms [16].

The aesthetic, discursive and materialistic components that have emerged from my work have as their unifying principle a conceptual framework that I have developed and termed Geometric and Genomic AbstractionISM, that is, geometric abstractionism guided by genomics and enabled by computer algorithms [17]. Like the art movements of cubism, created by Pablo Picasso and Georges Braque ca. 1907 [18], and simultanism (also called Orphism), created by Sonia and Robert Delaunay ca. 1912 [19]—movements that were self-contained artistic expressions characterized and distinguished by their own sets of techniques and methodologies—GAGAISMO also contains a concrete set of defined techniques and methods, derived from the convergence of genomics and bioinformatics with computational art.

In this article, I describe one such method that I conceived and developed in March 2018, which uses the nucleotide variation of genes from related grass species as raw material for the creation of artwork series that visually represent changes in DNA derived from plant speciation and evolutionary processes.

ORTHOGOLOGOUS GENE COPIES FROM RELATED GRASS SPECIES AS RAW MATERIAL FOR ART

Orthologous genes reflect the history of a species, diverging after a speciation event, and thus are related by vertical descent from a common ancestor [20]. This is certainly the case for the VERNALIZATION 1 (VRN1) gene copies from four different grasses that Higgens et al. had previously analyzed [21]. My interest in VRN1 gene copies originated in my art-based research on the occurrence and frequency distribution of CG and CCG sequence context variation across genic segments from VRN1 orthologous copies from Brachypodium dystachyon (purple false brome), Brachypodium stacei, Oryza sativa (rice) and Zea mays (maize) [22].

VRN1 is an intriguing component of a biological process that can be co-opted for art. VRN1 is responsible for conferring on a plant the ability to flower in response to extended periods of cold—known as the vernalization response—and thus provides a mechanism for molecular memory in plants [23]. Brachypodium dystachyon and Brachypodium stacei are plant species that respond to vernalization, flowering after long exposure to cold. Furthermore, their geographical distribution is associated with cold and temperate climates [24].
Rice and maize, in contrast, are vernalization-insensitive: They do not require exposure to cold to flower and they are grown in temperate and tropical regions of the world [25]. Although VRN1 genes have been demonstrated to have a functional role in the flowering response of both Brachypodium species, their role in rice and maize remains unknown. Nonetheless rice and maize harbor VRN1 gene copies [26]; thus it is possible to assume that VRN1 could have been under different selective pressure regimes in the Brachypodium species relative to rice and maize, and thus the genetic divergence at VRN1 gene copies among these plant species could be exploited for artistic purposes.

**METHODOLOGY**

Coding DNA sequences (CDS) for VRN1 genes were obtained from Phytozome [27] and aligned using MUSCLE [28]. The CDS used in this study can be accessed at the following links:

- Brast06g14030.1: https://phytozome.jgi.doe.gov/pz/portal.html?g=Bras&m=Brast06g14030.1&c=TIGR_BCAST06G14030.1
- Brast02g11100.1: https://phytozome.jgi.doe.gov/pz/portal.html?g=Bras&m=Brast02g11100.1&c=TIGR_BCAST02G11100.1
- Bradi1g08340.1: https://phytozome.jgi.doe.gov/pz/portal.html?g=Brad&m=Bradi1g08340.1
- LOC_Os03g54160.1: https://phytozome.jgi.doe.gov/pz/portal.html?g=Rice&m=LOC_Os03g54160.1
- GRMZM2G553379_T07: https://phytozome.jgi.doe.gov/pz/portal.html?g=Maize&m=GRMZM2G553379_T07
- GRMZM2G032339_T02: https://phytozome.jgi.doe.gov/pz/portal.html?g=Maize&m=GRMZM2G032339_T02

Since the total length of the sequence post-alignment was 852 base pairs, whereas the existing artworks selected and used as sources in the collage were 825 pixels wide, this meant that 96.8% of the VRN1 sequence post-alignment would be visually represented in the resulting artworks, with the remaining 27 base pairs from the 3’ end of the alignment (this terminology relates to DNA-RNA transcription) not considered for inclusion in the result and thus visually excluded from the resulting artwork. The resulting artworks represented almost the entire length of the VRN1 coding sequence and thus captured in visual format most of the DNA sequence variation between orthologous gene copies (Fig. 1).

**Fig. 1.** Multiple sequence alignment for CDS derived from VRN1 orthologous gene copies in Brachypodium stacei (Brast06g14030.1 and Brast02g11100.1), Brachypodium distachyon (Bradi1g08340.1), rice (LOC_Os03g54160.1) and maize (GRMZM2G553379_T07 and GRMZM2G032339_T02), respectively. Because Brachypodium stacei and maize underwent an additional round of whole genome duplication relative to Brachypodium distachyon and rice, respectively [35,36], they harbor two gene copies each (homeologous copies), constituting six CDS considered in this study instead of four. (© Martin Calvino)
CDS; thus it was practically convenient to use them as proof of concept. Also, the artworks selected are clear examples of geometric expressionism created with code and visually differ from one another, thus representing nucleotide divergence when collaged into new artworks after VRN1 CDS were aligned. Other artworks could have been used in the test case as well, as long as their dimensions along the x-axis would have been close enough to, or would have completely covered, the length of VRN1 CDS post-alignment.

For each VRN1 sequence taken from the alignment, I developed an algorithm, written in Processing [29], that assigned each nucleotide from the DNA sequence to a column of pixels in the existing artworks used as source (Color Plate A and online Supplemental Fig. 1).

For instance:

- nucleotide A: artwork #1 (Atardecer Geometrico by Martin Calvino, 2016)
- nucleotide G: artwork #2 (Etereo by Martin Calvino, 2016)
- nucleotide C: artwork #3 (Interferencia by Martin Calvino, 2016)
- nucleotide T: artwork #4 (Sequence by Martin Calvino, 2016)

The position of each column of pixels within the source artwork correlated with the position of its corresponding nucleotide in the sequence of DNA for VRN1 CDS post-alignment, giving a 1:1 correspondence. Because of this, the resulting artworks presented here could in principle be considered within the realm of data visualization in addition to computational art. Thus, I visually reconstructed VRN1 CDS post-alignment by arbitrarily assigning a column from artwork #1 each time there was an A on the DNA string, a column from artwork #2 each time there was a G on the DNA string, a column from artwork #3 each time there was a C on the DNA string, and a column from artwork #4 each time there was a T on the DNA string. I repeated this procedure for each of the six VRN1 CDS copies included in this study (Fig. 2). The correspondence between nucleotide variation among VRN1 CDS post-alignment and the resulting visual outcome of the technique described can be observed by comparing Fig. 1 with Fig. 2.

**SERIES OF VISUAL COLLAGES THAT REPRESENT THE EVOLUTIONARY HISTORY OF VRN1 ORTHOLOGOUS GENE COPIES**

The resulting artworks shown in Fig. 2 are visual reconstructions of each of the six VRN1 CDS post-alignment—in the form of a collage—built from columns of pixels each derived from one of the four artworks used as sources (Color Plate A). Differences among VRN1 CDS shown in Fig. 1 can be visually correlated with differences among the artwork series shown in Fig. 2, especially for three regions comprehended between 512 and 534 base pairs (region 1), 557 and 577 base pairs (region 2) and 601 and 645 base pairs (region 3), respectively.

**Fig. 2.** Visual outcome of reconstructing VRN1 CDS from Fig. 1 with artworks assigned to nucleotides from Color Plate A. Visual differences in artworks represent nucleotide divergence at VRN1 among plant species. From top to bottom: Brachypodium stacei (Brast06G240300.1 and Brast02G311100.1 artworks); Brachypodium distachyon (Bradi1g08340.1 artwork); rice (LOC_Os03g54160.1 artwork); and maize (GRMZM2G53379_T07 and GRMZM2G032339_T02 artworks). (© Martin Calvino)
Nucleotide differences toward the 3' end of the CDS from the homeologous copies of VRN1 in Brachypodium distachyon and the unique VRN1 gene copy in Brachypodium distachyon share the visual element derived from region 1 (312–354 base pairs) (Fig. 1 and first three artworks shown in Fig. 2), whereas rice VRN1 and maize homeologous VRN1 gene copies do not (Fig. 1 and last two artworks shown in Fig. 2). Although maize homeologous VRN1 gene copies share the visual element at region 3 (last two artworks shown in Fig. 3), differences in color among gene copies at this region are due to a G-to-A single nucleotide polymorphism (SNP) at position 600 base pair (Fig. 1). This implies that the evolutionary history of VRN1 genes in the grasses studied could be related with a visual story in the resulting artwork series, communicating a scientific finding in artistic form.

There are aesthetic differences between the resulting six artworks (Fig. 2) relative to the four artworks used as sources (Color Plate A). The visual heterogeneity derived from reconstructing VRN1 CDS post-alignment is not a priori the result of a random process but rather the result of evolutionary pressures shaping the nucleotide sequence of the genes. This evolutionary aspect needs to be emphasized, as the utilization of random functions in the generation of visual novelty is a popular approach among computational artists. A random process is not entirely at play here, since VRN1 gene copies can be argued to have been shaped by a combination of neutral evolution and natural selection [30]. For this reason, the aesthetics of visual elements from the source artworks were recontextualized under the framework of molecular evolution into something totally different. In my view, the "collaging agents" could be regarded as the genes themselves and the bioinformatic algorithm used to align them. The approach I developed, then, completely bypassed the need to use a random function in the generation of visual novelty when creating computational art.

**NUCLEOTIDE DIVERGENCE AMONG ORTHOLOGOUS GENE COPIES EXPLOITED FOR COMPUTATIONAL ART CREATION**

The approach I describe here demonstrates that nucleotide variation at genes derived from a common ancestor and presently conserved in related grass species can be co-opted as raw material to create a series of computational artworks. Furthermore, the visual outcome is true to data shown in the alignment and thus visually represents differences in nucleotide composition caused by the evolutionary divergence at the VRN1 gene in the four grass species I discuss.

The integration of plant genomics with computational art through the application of the method I describe can indeed be extended to other genes that share orthologous relationships among grasses, widening the prospect for art creation based on solid scientific concepts from the field of plant genomics and evolution. Furthermore, the work I present here can be extended to the alignment of protein sequences encoded from orthologous genes, as long as a unique artwork is assigned to each amino acid and the dimensions of the source artworks cover most of the protein sequences post-alignment.

Thus, nucleotide and protein divergence can be exploited as creative forces in the composition of artwork series, providing me with a valid methodology not only within my creative practice but also as a foundational technique in which to anchor the aesthetic, discursive and materialistic aspects of GAGAISMO. This approach could potentially evolve into an artistic movement as additional artists start to adopt the methodology presented and discussed here.

**USE OF DNA SEQUENCE FOR CREATION OF ART**

Genetic code has become the subject focus of many artists who have translated DNA sequences into visual forms or any other artistic mediums of expression. For instance, *Blind Genes* (2002) by Muller-Pohle translates DNA sequences from a database search for the keyword "blindness" into braille and creates digital paintings from those sequences [31]. Ben Fry's *Valence* (2002–2003), exhibited as part of the 2002 Whitney Biennial, visually represents the BLAST algorithm, a bioinformatic method for searching sequences in genome databases used worldwide by researchers in the life sciences [32]. *Genomixer* (2000–2005), by STANZA, uses DNA data taken from the artist himself to make visual artworks and sound-based systems [33]. More recently, Heather Dewey-Hagborg's *Stranger Visions* (2012–2013) uses mitochondrial DNA sequence data taken from chewed-up gum, cigarette butts and public bathrooms and waiting rooms across New York City, and algorithmically generates portraits that are then 3D-printed as life-size models [34].

Of these works, the one perhaps most relatable to my own is Fry's *Valence*. BLAST and multiple sequence alignments (MSAs) are essential algorithms within the field of genomics and bioinformatics, and Fry and I both co-opted these algorithms for computational art creation. It remains to be seen if all these independent efforts in co-opting tools for the conversion of DNA sequence data into art can be integrated (not only retrospectively, but also looking forward) into a single artistic discipline and/or movement; I would argue that GAGAISMO could be such a framework.

---

**References and Notes**


17 Hasdemir [9].


22 Calvino [14].


Manuscript received 18 July 2018.

MARTIN CALVINO is a multimedia artist and scientist. He was an artist-in-residence at the Art and Artificial Intelligence Lab at the Computer Science Department of Rutgers University. He was a senior research assistant at the School of Creative Media, City University of Hong Kong, and previously attended the Interactive Telecommunications Program (ITP) at Tisch School of the Arts, New York University. He holds degrees in the Interactive Telecommunications Program (ITP) at Tisch Media, City University of Hong Kong, and previously attended the Interactive Telecommunications Program (ITP) at Tisch School of the Arts, New York University. He holds degrees in plant molecular genetics from Rutgers University (PhD) and Tsukuba University (MS) and in molecular biology from the University of the Republic of Uruguay (BS).
Artworks used as source material for visual reconstruction of VRN1 CDS post-alignment shown in Fig. 1 on page [TK]. Each of four computational artworks (all with dimensions of 825 × 315 pixels) were assigned to each of the four nucleotides present in a string of DNA (A, G, C, T). (© Martin Calvino) (See the article in this issue by Martin Calvino.)