Bioinformatics Advance Access published August 12, 2004

Origins of introns based on the definition of exon modules and their conserved interfaces

Albert D.G. de Roos

albert.de.roos@thebeaglearmada.nl, +31.(0)6.14264218, The Beagle Armada, Postbus 964, 4600 AZ Bergen op Zoom, The Netherlands

Abstract

Central in the unraveling of the early evolution of the genome is the origin and role of introns. The evolution of the genome can be characterized by a continuous expansion of functional modules that occurs without interruption of existing processes. The design-by-contract methodology of software development offers a modular approach to design that seeks to increase flexibility by focusing on the design of constant interfaces between functional modules. Here, it is shown that design-by-contract can offer a framework for genome evolution. The definition of an ancient exon module with identical splice sites leads to a relatively simple sequence of events that explains the role of introns, intron phase differences and the evolution of multi-exon proteins in an RNA world. An interaction of the experimentally-defined six-nucleotide splicing consensus sequence together with a limited number of primitive ribozymes can account for a rapid creation of protein diversity.

Introduction

One of the most intriguing questions in unravelling genomic evolution is whether the intron/exon structure of eukaryotic genes reflects their ancient assembly by exon shuffling or whether the introns have been inserted into preformed genes. Several theories have been put forward to explain the role of introns and exons in evolution (reviewed in Mattick, 1994; Logsdon, 1998; Fedorova and Fedorov, 2003; Rzhetsky and Ayala, 1999). There are now two main competing theories that try to explain the role of introns and are both based on the involvement of DNA based introns and exons . The ‘introns early’ or exon theory of genes states that the introns are ancient and have been subsequently lost in prokaryotes (Gilbert, 1987; Gilbert et al, 1996; Gilbert et al, 1997). In this theory, the first exons coded for ancient protein modules from which multi-modular proteins were assembled by means of exon shuffling and recombination. Introns facilitated this process by providing the actual sites of recombination. On the other hand, the ‘introns late theory’ maintains that the spliceosomal introns were inserted into the eukaryote genes later in evolution (Palmer and Logsdon, 1998; Cavalier-Smith, 1991; Cho and Doolittle, 1997; Logsdon, 1998) after the evolution of multi-modular proteins. In introns-late, the appearance of introns could also have aided in the creation of diversity by facilitating recombination. No conclusive evidence has been found to prove or disprove intron-early or intron-late,

although these theories are based on completely different genome architectures and mechanisms of evolution.

The genome has evolved from a simple RNA based self-replicating system, the RNA world (Gilbert, 1986; Joyce, 2002) to a complex system of multi-exon genes coding for multi-modular proteins. During this evolutionary process, numerous new functions were added or modified without disrupting the functioning of older systems. The evolution from strands of RNA to multi-exon genes with sophisticated expression systems implies that the genome was able to increase in size and complexity many orders of magnitude without losing flexibility. Any genome architecture that would form the basis of genome evolution should therefore be flexible and robust in order to meet the requirements for virtual unlimited expansion of size and function.

Modern software designs seek to increase flexibility by using a modular approach which allows for the addition, replacement and changing operations within individual modules. Complex software architectures are based on a methodology in which a software system is viewed as a set of communicating modules whose interaction is based on precisely defined interfaces. The interfaces can be viewed as specifications of the mutual obligations or *contracts*. The effect of constant interfaces across modules is a reduction of the interdependencies across modules or components and a reduction of the risk that changes within one module will create unanticipated changes in other modules. This methodology is also known as design-by-contract (Meyer, 1997). Since the characteristics of the design-by-contract methodology are similar to those required in genome evolution, it is hypothesized here that genome architecture reflects the paradigms of design-by-contract: definition of functional modules that interact with each other by well-defined interfaces.

Modularity and interfaces in the genome

The basic unit of genetic information, the gene, can be regarded as a self-contained module with a well-defined interface. A gene contains all the necessary information from which the encoded protein can be generated, whereas the highly conserved genetic code functions as the interface between gene and protein. Eukaryotic genes consist themselves of parts of coding sequences, exons, interrupted by non-coding sequence, the introns (**Fig. 1A**). The introns have to be spliced out in order to form a continuous coding sequence, mRNA, that can be recognized by the translation machinery. In principle, an intron contains all the necessary information to be spliced out which enables it to function independently from the exon sequence. The intron can therefore be regarded as a self-contained modules with a well-defined (conserved) interface, the splice recognition site (**Fig. 1B**), which is located exclusively in the intron. This configuration enables the excision of introns independent from exon sequence.

Exons are, in contrast to introns, dependent upon information that lies outside of the exons, since the splice recognition sites of the intron determine the span of the exon. A dependence on intron sequences would severely hamper independent movement and exchange of coding sequences between genes. However, extensive recombination of exons by exon shuffling is believed to played an important role

in the creation of genetic diversity (Patthy, 2003; Sudhof *et al.*, 1985; Kolkman and Stemmer, 2001) and many of the proteins with functionally divergent domains were established before the division of prokaryotes from eukaryotes (Ohno, 1987). In order to be inserted into random nucleotide sequences, the exon module should preferably behave like a self-contained module. The exon would largely acquire independence when the conserved intron sequences that flank both ends of the exon would be included as part of the exon (**Fig. 1C**), enabling it to function as an independent coding module, or ancient exon module.

Molecular view on the ancient exon

The ends of the proposed ancient exon module were studied in more detail at a molecular level using an intron-exon database (Clark, 2003) generated from Genbank release 127 (Benson et al, 2002). The last nucleotides on either side of the exon module are represented by the intron-exon boundary and possible remnants of a consensus sequence were determined by looking at nucleotide triplets from the intron and the exon part of the intron-exon boundary. The tri-nucleotide sequences with the highest frequencies of several species are shown in **Table 1**. Looking at the overall similarity between the sequences on both ends of the exon module and the conservation of these sequences between species, a bias towards the sequence CAG|GTG can be discerned both in the sequence preceding the exon as in the one following the exon. No significant differences were observed between sequences from intron-exon boundaries with different intron phases (results not shown).

Based on the data in table 1, it is proposed that the conserved sequences of both ends of this ancient functioned as the ancient exon recognition site with an original sequence CAGGUG (**Fig. 2A**). This consensus sequence could have served as cleavage recognition site enabling the splicing out of the coding sequence, creating the substrate for the translation machinery (**Fig. 2B**). A cleavage in the middle of the sequence CAGGUG would result in a spliced out coding RNA sequence that is always surrounded by the remaining parts of the recognition sequence, GUG at the start and CAG at the end of the exon. Concatenation of these ancient exon modules after cleavage of the recognition sites would join the remaining parts of the recognition sequence (**Fig. 2C)**, forming multi-exon mRNA.

Support for the existence of the ancient splice site can be provided by the fact that the codon GUG still acts as a translation start site in bacteria (Gold, 1988) and can still function as one in other organisms (Mehdi *et al.*, 1990; Peabody, 1989). Moreover, the most common start codon AUG differs only one nucleotide from GUG and a single mutation of the first nucleotide of the hypothetical ancient end sequence CAG is needed to convert it into the *amber* stop codon UAG. Other support for the role of the ancient splice site comes from the intron-less genes of prokaryotes. It has been shown that the coding sequences around the positions of introns insertion in their eukaryotic counterparts also show a consensus sequence CAG^GT, originally dubbed the proto-splice site (Dibb and Newman, 1989). If introns were lost during evolution in an RNA world with a mechanism closely related to splicing (cf. Fig. 2C), the proposed ancient splice site would also be retained.

Exon phase and frame-shift

The joining of two exons modules as shown in figure 2C implies that part of the consensus splicing sites become part of the coding sequence (**Fig. 3A**) and every module would be connected by a fixed series of 6 nucleotides, formed by the sequence CAGGUG. The two codons in this sequence (CAG and GUG) would always be translated into the amino acids glutamine (Q) and valine (V). In our design-by-contract model, the recognition sequence represents the interface for the splicing of the exons and therefore, any mutation in this sequence would be deleterious since it would result in the inactivation of the splice site (**Fig. 3B**) and resulting loss of function of the encoded protein. On the other hand, mutation of the amino acid sequence would be evolutionary advantageous since it would relieve the obligatory translation of the ancient splice site into the amino acids Q and V. A phase shift would enable the reading-through of the recognition sequence in another way (**Fig. 3C**), leading to a different amino acid sequence between the exons with an identical recognition consensus sequence.

The actual distribution of amino acids at splice junctions was investigated using an exon-intron database containing phase information (Sakharkar *et al.*, 2000) derived from Genbank 122 (Benson *et al.*, 2002). **Figure 4** shows that the last amino acid of an exon has a phase-dependent preference for specific amino acids. In each phase, the last amino acid follows closely the ones that can be predicted from a phase shift based on a constant splice recognition sequence (Fig. 3C). Note that at the nucleotide level, the intron-exon boundary does not exhibit phase-dependent differences (not shown). **Table 2** shows that the amino acid positions that would be have been affected by an ancient phase shift still show a bias towards their predicted phase. This effect is even stronger when the effect of a phase shift is viewed in both exons simultaneously, up to the point that almost 95% of the amino acid sequence Q|V around a splice site is in phase 0.

Since splicing out of introns is necessary for correct translation, intronless mRNA can be considered as a well-conserved interface to the translation machinery. The generation of intronless mRNA by a concatenation of different coding RNA modules in random ‘intron’ RNA sequence (Fig. 2), would not change this interface and could take place without affecting translation. Also, the separate development of functional protein modules, followed by an assembly of these modules would be inherently less complex and more flexible (Gilbert, 1987; Patthy 2003). Phase shift could be viewed as an outcome of a genomic evolution model based on the design-by-contract methodology, since phase shifts could provide a means for creating more protein diversity without affecting the established splicing interface. The development of a splicing machinery that would confine the splicing recognition sequence exclusively to the intron (as is presently the case, cf. Figure 1B) would ultimately enable the complete independent evolution of the coding ends of the exon.

The degree of conservation at the boundaries of exons flanking introns has been shown before and has been interpreted as a derived result of evolution for efficient splicing (Long *et al.*, 1997), the preferred insertion site for introns (Dibb and Newman, 1989) or as functional splice sites that existed in the coding sequence of genes prior to the insertion of introns (Sadusky *et al.*, 2004). Intron phase has been

shown to be correlated to codon position (Long *et al.*, 1995; Tomita *et al.*, 1996) and hypothesized to be related to exon shuffling between exons in the same phase (Long *et al.*, 1995).

Fundamental steps in evolution based on a single template

It is proposed here that the sequence CAGGUG acted as the ancient cleavage recognition site for a ribozyme. Ribozymes can interact with its targets b a complementary RNA sequence primarily based on Watson-Crick base pairing (Guerrier-Takada *et al.*, 1989; Cech, 1987). Based on the sequence of the ancient splice site, an antiparallel arrangement of this sequence could interact with itself (**Fig. 5A**), making a single recognition sequence act as both the target site and the target recognition sequence. At a molecular level, this interaction could be stabilized by four pairs of Watson-Crick base pairs while leaving two G-pairs unpaired.

The splicing out of the RNA sequences between the exon modules, equivalent to intron splicing, is an important step in genome evolution. **Figure 5B** shows how the antiparallel arrangement of two adjacent exon modules could facilitate splicing. In addition to an intra-strand cleavage between G residues, a religation of the G's to the opposing strand would concatenate the two exons, a process that could be facilitated by the close physical proximity of the G's involved.

Another important step in the evolution of proteins is the exchange of coding sequences between different genes resulting in the recombination of genes. A mechanism identical to intron splicing as shown in figure 5B but followed by an *in trans* religation would lead to the exchange of RNA strands (**Fig. 5C**) between RNA molecules. In this way, ancient ribozymes could have played an active role in the generation of the diversity of proteins.

Thus, based on a six-nucleotide proto-splice site and relatively simple ribozymes that could cleave and religate this sequence, three important events in the exon-centric evolution of multi-domain proteins can be explained: i) the splicing out of the exon modules yielding short exonic mRNA, ii) the splicing out of RNA sequences between exon thereby concatenating exon modules to multi-exon mRNA, and iii) the active recombination of exons. The classes of ribozymes that could catalyse the cleavage and ligation reactions proposed in figure 5 have been shown to occur naturally (Symons, 1992; Guerrier-Takada, 1983). Ribozyme RNAse T1 cleaves a double stranded complementary RNA sequence at unpaired G residues), and apart from several naturally occurring RNA ligases ((Yoshida, 2001; Hager *et al.*, 1996), it has also been shown that complex ligases can evolve from group I ribozyme domains (Jaeger *et al.*, 1999) and from small random RNA sequences (Ekland *et al.*, 1995).

The proto-splice site can act as a starting point for the evolution of multifunction proteins when the consensus sequence of the proposed proto-splice site arises randomly in strands of RNA. Two splice sites in close proximity could then lead to the first functional single-exon genes. The transformation of the coding parts of the proto-splice site sequences into start (GUG to AUG) and stop codons (CAG to

UAG) and *vice versa* back to a functional proto-splice site could facilitate a stepwise concatenation of exons (cf. Fig 2).

The introns that arose early in evolution as a consequence of a concatenation of exons (Fig. 2) could be lost further in evolution, but their presence at conserved positions would still reflect their ancient origins. The evolution of transposons from introns, both able to function as relatively independent functional units, may account for many of the observations attributed to the introns-late theory (Cho and Doolittle, 1997; Logsdon, 1998).

Evolution on a design-by-contract theory

The application of the design-by-contract methodology by viewing the exon as a module that interacts with its environment by its interface, led to a series of logical steps explaining the intron-exon structure of genes and intron phase differences. It suggests that evolution behaved according to a design pattern that separates functional modules from each other by well-defined interfaces. The dependence of vital functions on interfaces prevent changes in the interfaces and force evolution in an architecture that reflects design-by-contract 'rules'. It also proposes that the major events leading to a diversification of proteins were situated in an RNA world. The next fundamental step in genome evolution, the transition from the RNA world to the RNA/DNA world can also be explained in line with design-by-contract. In order to keep al the interfaces that were created in the RNA world intact, the entire RNA genome could have been copied *verbatim* into DNA.

References

Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Rapp B.A., Wheeler D.L. (2002). GenBank. Nucleic Acids Res. 30: 17-20.

Cavalier-Smith, T. (1991) Intron phylogeny: a new hypothesis. Trends Genet. 7, 145-148

Cech T.R. (1987). The chemistry of self-splicing RNA and RNA enzymes. Science 236: 1532-9.

Cho, G. & Doolittle, R. F. (1997) Intron distribution in ancient paralogs supports random insertion and not random loss. J. Mol. Evol. 44, 573-584

Clark, F. (2003) Gene data sets derived from GenBank [web page]; http://www.maths.-uq.edu.au/~fc/datasets/.

Dibb N.J., Newman A.J. (1989). Evidence that introns arose at proto-splice sites. EMBO J. 8: 2015-21.

Ekland E.H., Szostak J.W., Bartel D.P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. Science 269: 364-70

Fedorova L., Fedorov A. (2003) Introns in gene evolution. Genetica. Jul;118(2-3):123-31.

Gilbert, W. (1986) The RNA World. Nature. 319: 618.

Gilbert, W. (1987) The exon theory of genes. Cold Spring Harb. Symp. Quant. Biol., 52: 901-5.

Gilbert, W., Marchionni, M. & McKnight, G. (1986) On the antiquity of introns. Cell. 46: 151-3.

Gilbert, W., de Souza, S. J. & Long, M. (1997) Origin of genes. Proc. Natl. Acad. Sci. U S A. 94: 7698-703.

Gold L. (1988). Posttranscriptional regulatory mechanisms in Escherichia coli. Annu. Rev. Biochem. 57: 199-233.

Guerrier-Takada C., Gardiner K., Marsh T., Pace N., Altman S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell. 35: 849-57.

Guerrier-Takada C., Lumelsky N., Altman S. (1989). Specific interactions in RNA enzyme-substrate complexes. Science 246: 1578-84.

Hager A.J., Pollard J.D., Szostak J.W. (1996) Ribozymes: aiming at RNA replication and protein synthesis. Chem. Biol. 3: 717-25.

Jaeger L., Wright M.C., Joyce G.F. (1999) A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. Proc. Natl. Acad. Sci. U S A. 96: 4712-7.

Joyce GF (2002) . The antiquity of RNA-based evolution. Nature. 418:214-21.

Kolkman J.A., Stemmer, W.P. (2001) Directed evolution of proteins by exon shuffling. Nat. Biotechnol. 19:423-8.

Logsdon, J. M. (1998) The recent origins of spliceosomal introns revisited. Curr. Opin. Genet. Dev. 8, 637-648.

Long M., de Souza S.J., Gilbert W. (1997) The yeast splice site revisited: new exon consensus from genomic analysis. Cell. 12:739-40.

Long M., Rosenberg C., Gilbert W. (1995) Intron phase correlations and the evolution of the intron/exon structure of genes. Proc. Natl. Acad. Sci. U.S.A. 92:12495-9.

Mattick, J.S. (1994). Introns: evolution and function. Curr. Opin. Genet. Dev. 4:823-31.

Mehdi H., Ono E., Gupta K.C. (1990). Initiation of translation at CUG, GUG, and ACG codons in mammalian cells. Gene. 91:173-8.

Meyer B. (1997) Object-Oriented Software Construction, Second Edition, Prentice Hall

Ohno S. (1987) Early genes that were oligomeric repeats generated a number of divergent domains on their own. Proc. Natl. Acad. Sci. USA. 84: 6486-6490

Palmer, J. D. & Logsdon, J. M. (1991) The recent origins of introns. Curr. Opin. Genet. Dev. 1, 470-477

Patthy L. (2003) Modular assembly of genes and the evolution of new functions. Genetica. 118: 217-31.

Peabody D.S. (1989). Translation initiation at non-AUG triplets in mammalian cells. J. Biol. Chem. 264: 5031-5.

Rzhetsky, A., Ayala, F.J. (1999) The enigma of intron origins. Cell. Mol. Life Sci. 55: 3-6

Sudhof, T.C. Goldstein, J.L., Brown, M.S. and Russell, D.W. (1985) The LDL receptor gene: a mosaic of exons shared with different proteins. Science, 228, 815–822.

Sakharkar M., Long M., Tan T.W., de Souza S.J. (2000). ExInt: an Exon/Intron database. Nucleic Acids Res. 28: 191-192.

Sadusky T., Newman A.J., Dibb N.J. (2004) Exon junction sequences as cryptic splice sites: implications for intron origin. Curr. Biol. 14:505-9.

Symons R.H. (1992). Small catalytic RNAs. Annu. Rev. Biochem. 1992;61:641-71

Tomita M., Shimizu N., Brutlag D.L. (1996) Introns and reading frames: correlation between splicing sites and their codon positions. Mol Biol Evol. 13: 1219-23.

Yoshida H. (2001). The ribonuclease T1 family. Methods Enzymol.341:28-41.

Table headers and Figure legends

Table 1. Frequency distribution of the nucleotide sequences of the intron-exon boundary at the ends of the putative ancient exon module. **A**. The left hand side or 5' end of the exon module, consisting of the sequence of the last three RNA nucleotides of the intron (3' intron) followed by the first three nucleotides of the exon (5' exon). **B**. The right hand side or 3' end of the exon module consisting of the 3' end of exon and the 5' end of the intron. Only the first 5 most abundant tri-nucleotide sequences are shown. The 5' ends of the first exon of a gene and the 3' part of the last exon of a gene were not included in the data. Controls (ctrl) show nucleotide sequences in exons and introns 12 nucleotides up- or downstream from the splice site. **Methods**. Data sets with intron and exon data were downloaded from the internet page of Francis Clark (Clark, 2003). These data are based on gene annotation in DNA sequences derived from Genbank (Benson *et al.*, 2002). The relevant intron and exon data were extracted from these files and converted to a tab-limited text file that was imported into tables created in a MySQL database (www.mysql.com). Exon tables consisted of a unique gene identifier, the exon number and exon sequence, and the intron tables of the gene identifier, intron number and intron sequence. A third table consisted of the gene identifiers with their corresponding Genbank accession number enabled joining with other Genbank databases. SQL queries are available upon request.

Table 2. Relationship between amino acids bordering the splice site and intron phase. Intron phase distribution with indicated amino acid preceding or following a splice site, and a combination of the amino acids. Amino acid chosen reflect the proposed mechanism of intron phase differences in Fig. 3. Controls are phase distributions in exons without the indicated sequences in that row. **Methods**: see Fig. 4.

Fig. 1. From an intron-centric to an exon-centric view on the structure of eukaryotic genes. **A**. Generalized eukaryotic gene structure. Eukaryotic genes consist at the RNA level of coding sequences (exons) interspersed with non-coding sequences (introns). Before translation, the introns are spliced out to form a continuous coding sequence, the messenger RNA (mRNA). **B.** The intron in detail. Introns are spliced out based on conserved sequences at both ends of the intron. At the left-hand side the three most common ribonucleotides sequence is GUG, at the right hand side this sequence is CAG. The bold characters indicate residues necessary for splicing. **C**. The ancient exon viewed as a unit. In an exon-centric view on the gene structure, the conserved parts of the intron could have functioned as signals demarcating the end and beginning of the exon.

Fig. 2. An exon-centric view on gene structure. **A**. Eukaryotic exon showing the generalized sequences at the intron-exon and exon-intron boundaries. The most common exon sequence on the 5' of the exon is GTG, on the 3' side it is CAG. Both sides of the exon now have an identical 6 ribonucleotide long sequence. **B**. The exon unit can be viewed as a coding sequence, surrounded by 2 identical recognition

sequences (CAGGUG), where actual splicing could occur in the middle of the recognition site, the exon-intron boundary. **C**. Concatenation of exons based on a single recognition site would join exons at their splice sites while retaining parts of the recognition sequences in the resulting mRNA at the splice junctions. This sequence is identical to the original recognition site and could be the equivalent of the proposed proto-splice sites in prokaryotic intron-less genes. Note the recurrent sequence CAGGUG.

Fig. 3. mRNA intron phase shifts can change amino acid sequence without changing the ancient splice recognition site. **A**. The sequence at the splice junctions codes for the amino acids glutamine and valine that would result from concatenation of the exons **B**. Mutations in the splice site recognition sequence will disrupt splicing. **C**. Phase shifts can change the splice junction amino acid sequence without disruption of the recognition site. In phase 0, there is only one amino acid sequence possible, glutamine followed by valine. In phase 1 and 2 translation will read through the splice junction in a different way, making various combinations of amino acids possible. Note the conservation of the sequence at the splice junctions (in red).

Fig. 4. Amino acid frequencies around splice sites in different phases. The last amino acid coded by each exon was determined when the exon was followed by a phase 0 (**A**), phase 1 (**B**) and phase 2 intron (**C**). In phase 0, the last amino acid just before the splice site is shown, in phase 1 and 2 the amino acid was taken that bridged the actual splice site on a nucleotide basis. As a control, amino acids three residues downstream the splice site were taken (cf. Fig. 3C). The control values were similar in each phase and were therefore averaged. In general, the amino acid distribution follows the nucleotide triplet data presented in table 1, although some differences can be seen due to the fact that amino acids can have multiple codons and that in figure 4 all species present in Genbank are pooled. **Methods**. The ExInt database (**Error! Bookmark not defined.**) containing a set of tables with exon and intron data including exon amino acid sequence, intron phase data and Genbank accession number was kindly provided by dr. Meena Sakharkar. SQL queries were performed that determined the last amino acid that was coded by the respective exon in each phase.

Fig. 5. Fundamental events in the evolution of multifunction, multi-exonic proteins based on a single recognition sequence. **A**. Cleavage. The sequence CAGGUG functions as the signal sequence for cleavage while the actual recognition takes place via the same sequence of a ribozyme using canonical (GC and UA) base pairing, possibly combined with non-canonical GG pairing. **B**. Cleavage and *in cis* religation. On the basis of an anti-parallel arrangement of recognition sites, splicing could be accomplished by a cleavage of the two recognition sites followed by a religation between opposing strands. **C**. Cleavage and i*n trans* religation. A religation between different RNA strands leads to a recombination of exons modules.

Tables and Figures: Table 1A

	intron 3'			ctrl	exon 5'			ctrl
Human	**cag**	16304	68.7	**2.3**	**gtg**	1767	7.4	**1.7**
	tag	5272	22.2	**1.0**	gag	856	3.6	**2.2**
	aag	1108	4.7	**1.6**	gaa	841	3.5	**2.4**
	gag	129	0.5	**2.3**	ggt	818	3.4	**1.7**
	agg	46	0.2	**2.6**	gtt	814	3.4	**1.2**
Drosophila	**cag**	25565	64.1	**1.0**	**gtg**	1465	3.7	**1.4**
	tag	10063	25.2	**1.1**	atg	1456	3.6	**1.5**
	aag	2409	6.0	**1.5**	atc	1370	3.4	**1.6**
	gag	233	0.6	**0.8**	att	1335	3.3	**1.7**
	taa	76	0.2	**3.5**	aaa	1197	3.0	**2.6**
Arabidopsis	**cag**	56151	60.8	**0.7**	gtt	7033	7.9	**1.9**
	tag	26371	28.5	**1.2**	**gtg**	5145	5.8	**1.0**
	aag	6703	7.3	**1.4**	gaa	4295	4.8	**2.9**
	gag	1228	1.3	**0.8**	gta	4071	4.6	**0.7**
	tga	140	0.2	**2.6**	gat	3505	3.9	**2.2**
Caenorh.	**cag**	76340	81.5	**0.8**	aaa	5315	5.7	**2.7**
	tag	12994	13.9	**1.2**	att	4901	5.2	**2.0**
	aag	3347	3.6	**1.6**	aat	3549	3.8	**2.0**
	gag	578	0.6	**0.7**	gaa	3509	3.7	**2.9**
	ttt	29	0.0	**7.7**	atg	3404	3.6	**1.6**
		#	%	%		#	%	%

Tables and Figures: Table 1B

	exon 3'			*ctrl*	intron 5'			*ctrl*
Human	**cag**	5566	23,5	**2.8**	gta	11817	49,7	**0.8**
	aag	3990	16,9	**1.6**	**gtg**	10073	42,4	**2.2**
	gag	2245	9,5	**2.2**	gtc	720	3,0	**1.3**
	ctg	1003	4,2	**3.0**	gtt	553	2,3	**1.3**
	atg	842	3,6	**1.5**	gca	113	0,5	**1.5**
Drosophila	aag	4866	12,2	**1.7**	gta	21804	54,5	**1.3**
	cag	4591	11,5	**1.9**	**gtg**	14279	35,7	**0.9**
	gag	2553	6,4	**1.6**	gtt	2164	5,4	**1.7**
	atg	1578	3,9	**1.5**	gtc	749	1,9	**0.7**
	caa	1468	3,7	**3.0**	gca	127	0,3	**1.2**
Arabidopsis	aag	16985	19,1	**2.3**	gta	58785	63,6	**1.2**
	cag	15845	17,8	**1.3**	gtt	16100	17,4	**0.7**
	gag	9348	10,5	**1.8**	**gtg**	11588	12,5	**0.8**
	atg	4114	4,6	**1.6**	gtc	4698	5,1	**2.0**
	ctg	3482	3,9	**1.1**	gca	439	0,5	**1.3**
Caenorh.	aag	11184	12,0	**1.8**	gta	51869	55,3	**1.1**
	cag	9175	9,9	**1.3**	**gtg**	23717	25,3	**0.7**
	gag	6709	7,2	**1.3**	gtt	16020	17,1	**1.8**
	aaa	4923	5,3	**3.8**	gtc	1526	1,6	**0.5**
	atg	4149	4,5	**1.8**	gca	179	0,2	**0.8**
		#	%	**%**		#	%	**%**

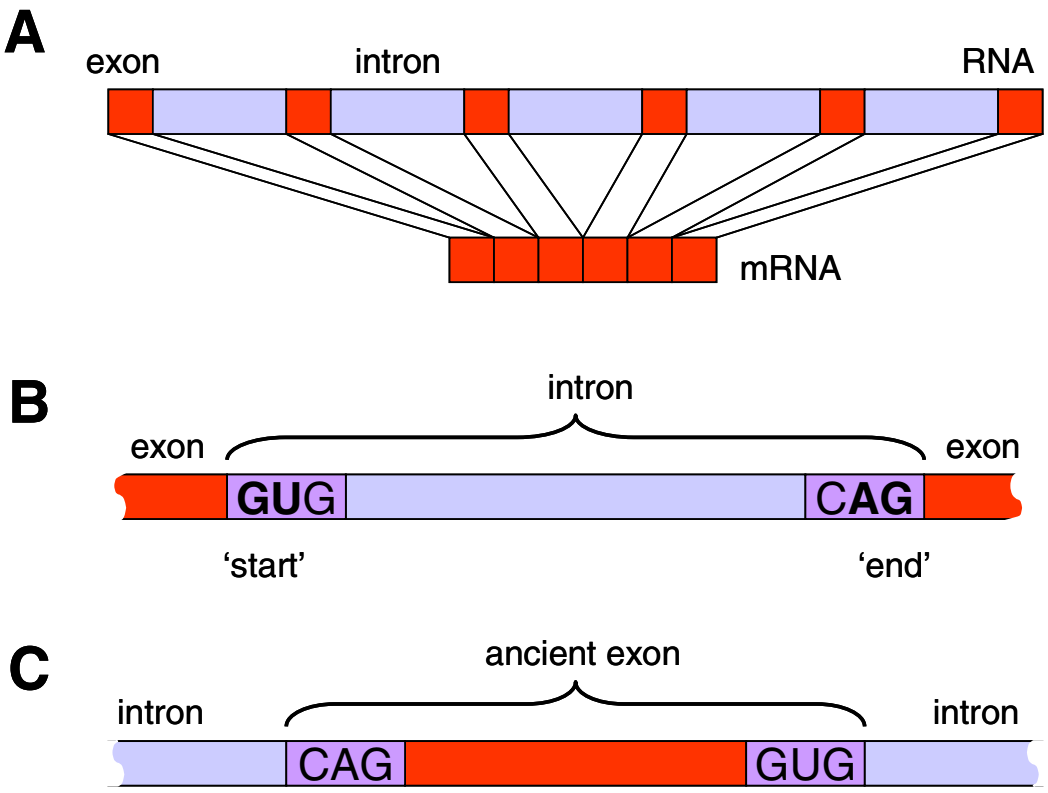
Tables and Figures: Table 2

		pre Q		post V		pre Q, post V		control	
phase	**0**	46544	91.5	28121	63.1	5688	94.8	188735	43.2
	1	1717	3.4	8895	20.0	132	2.2	137145	31.4
	2	2634	5.2	7537	16.9	180	3.0	110541	25.3

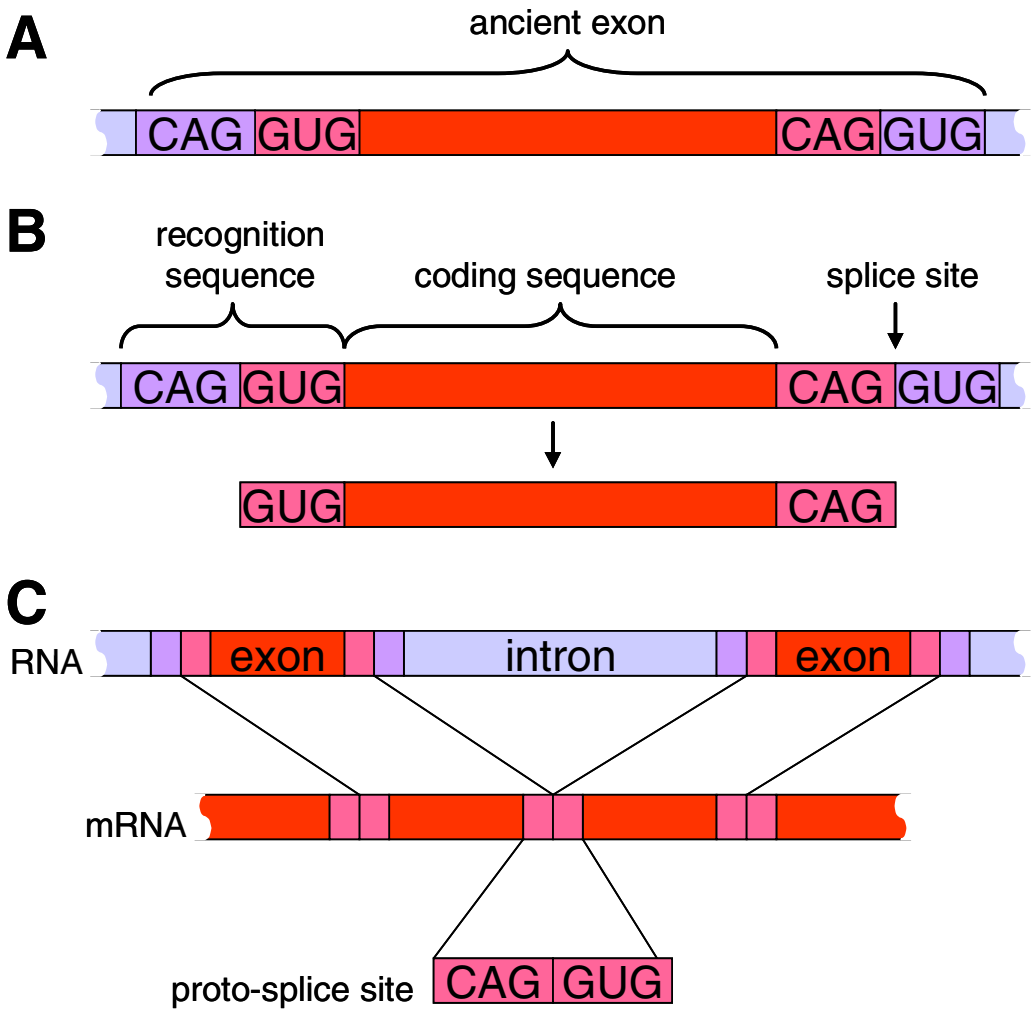
		pre-1 TPAS		pre G		TPAS/G		control	
phase	0	57490	43.9	5080	8.5	1411	7.1	196553	55.4
	1	45451	34.7	44377	73.8	15757	79.2	73554	20.7
	2	28116	21.5	10648	17.7	2716	13.7	84484	23.8

		pre R		post CW		R+CW		control	
phase	0	11243	23.3	6071	37.4	283	16.0	240681	52.0
	1	3909	8.1	5437	33.5	158	8.9	138437	29.9
	2	32999	68.5	4732	29.1	1328	75.1	84129	18.2
		#	*%*	*#*	*%*	*#*	*%*	*#*	*%*

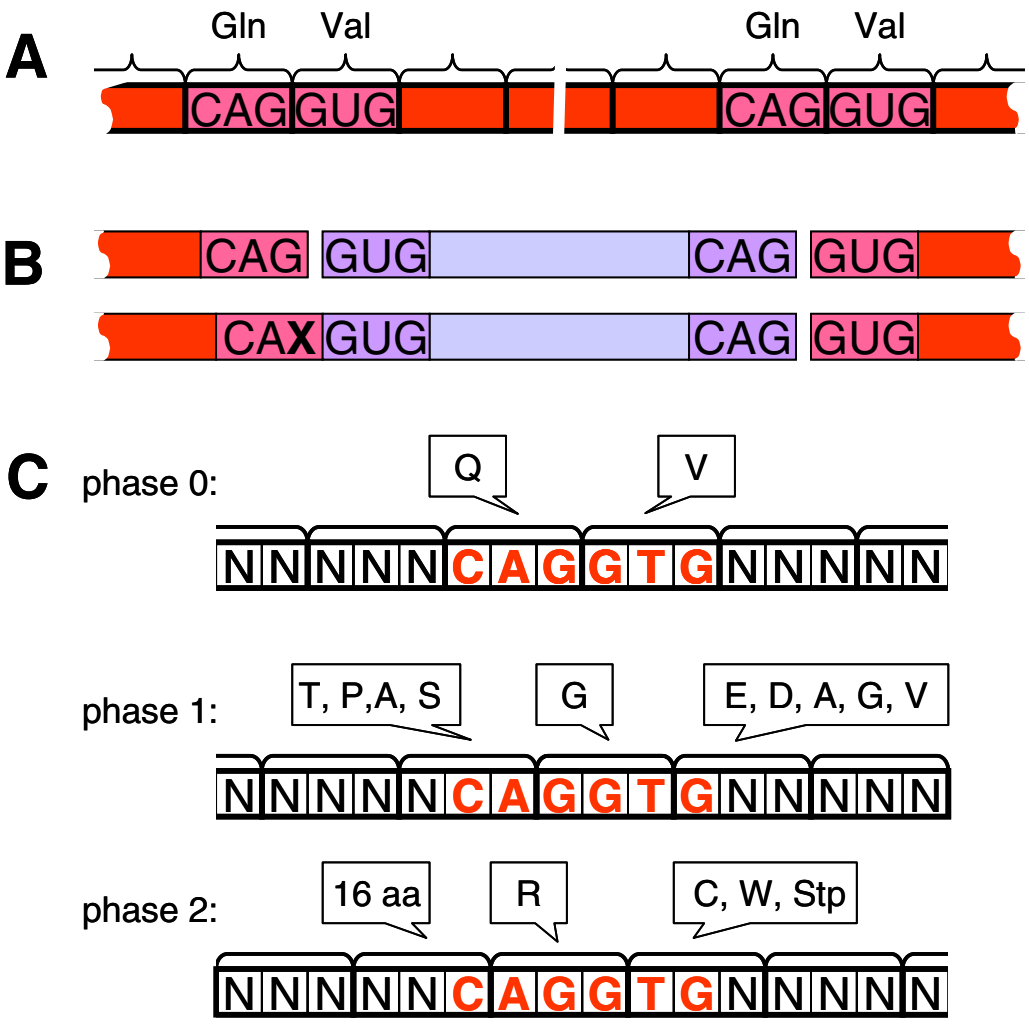
Tables and Figures: Figure 1



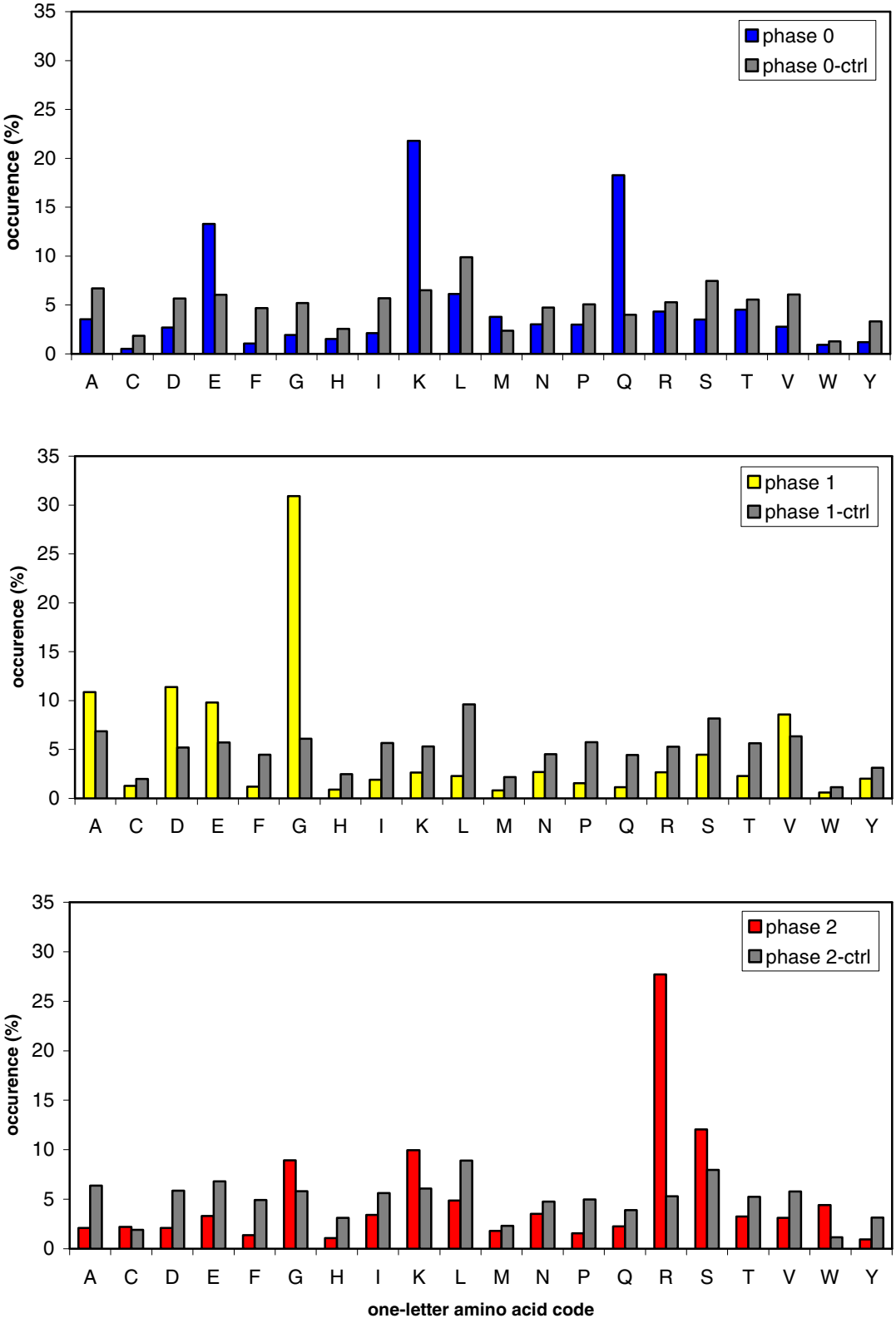
Tables and Figures: Figure 2



Tables and Figures: Figure 3



Tables and Figures: Figure 4



Tables and Figures: Figure 5