

PART TWO

Table SMO. Overview of the sequence_features and sequence_attributes (based on the Sequence Ontology project) identified by the GenomeContent from the gene annotation files (GTF* or GFF*) of 461 eukaryotic genome projects

Lozada-Chávez, I., Stadler, P.F., and Prohaska, S.J. (2018). bioRxiv,doi.org/10.1101/283549. Re-submitted (2022)
Genome-wide features of introns are evolutionary decoupled among themselves and from genome size throughout Eukarya
<https://www.biorxiv.org/content/early/2018/03/18/283549>

Sequence Ontology project: <https://github.com/The-Sequence-Ontology/SO-Ontologies>

Official SO sequence features (e.g., "gene", "CDS", "mRNA") and attributes (e.g., "Parent", "gene_id", "ID", "name") in columns 2, 3 and 9 are recognized from the **SOFA.obo** file:

https://github.com/The-Sequence-Ontology/SO-Ontologies/blob/master/Ontology_Files/subsets/SOFA.obo

GFF and GTF sequence features from genome projects														
species	taxon	id	type	name	version	url	assembly	chromosome	length	start	end	strand	feature	
picoides_pubescens	metazoa	485931	genbank	GFF-like	gene : mRNA : CDS	RefSeq Gnomon iRNAscan-SE			3	ncRNA18 : mRNAs16 : CDS : gene : mRNA_segment 1 ; transcript 6 : region 31254 iRNA 239 ; exon 175977 CDS 178802 gene 14476 V_gene_segment 3 ; cDNA_match 11815				
pileoderma_croceum	fungi	240852	jgi	GTF-like	gene : CDS	JGI			1	CDS 97851 exon 10258 stop_codon 20418 start_codon 20055				
pinctada_fucata	metazoa	887626	oistjp	GFF-like	gene : transcript : CDS	AUGUSTUS			2	stop_codon 31547 transcription_start_site 2396 gene 33306 transcript 35066 CDS 23328 start_codon 30725 transcription_end_site 31204 exon 25148				
pinus_taeda	plantae	528596	pinegenome	GFF-like	gene : mRNA : CDS	maker			1	mRNA 50172 three_prime_UTR 16450 gene 50172 exon 197597 live_prime_UTR 23049 CDS 191156				
pinorphysora_indica	fungi	144528	jgi	GTF-like	gene : CDS	JGI			1	stop_codon 11554 start_codon 11604 exon 60685 CDS 60685				
pirimyces_sp	fungi	112453	jgi	GTF-like	gene : CDS	JGI			1	stop_codon 145081 stop_codon 12275 exon 10529				
pisolithus_microcarpus	fungi	202448	jgi	GTF-like	gene : CDS	JGI			1	CDS 76774 start_codon 11599 stop_codon 11728 exon 7728				
pisolithus_innotatus	fungi	222051	jgi	GTF-like	gene : CDS	JGI			1	stop_codon 20334 CDS 8886 start_codon 1662 exon 25148				
plasmiodiphora_brasiccae	protists	163628	ensembl	GFF-like	gene : transcript : CDS	ena_exon ena ena_assembly_gap			3	three_prime_UTR 666 : gene 9727 exon 44646 transcript 9727 five_prime_UTR 3 CDS 44646 biological_region 45199				
plasmodium_berghei	protists	35414	ensembl	GTF-like	gene : transcript : CDS	rRNA pseudogene ncRNA protein_coding iRNA			5	exon 12947 stop_codon 4838 start_codon 4839 CDS 12790				
plasmodium_falciparum	protists	40355	ensembl	GTF-like	gene : transcript : CDS	protein_coding rRNA non_translating_CDS snRNA iRNA ncRNA pseudogene			6	CDS 14315 exon 15101 start_codon 5471 stop_codon 5468				
plasmodium_vivax	protists	35387	ensembl	GTF-like	gene : transcript : CDS	rRNA_p_MRP RNA snoRNA snRNA iRNA protein_coding			9	start_codon 5050 exon 12691 stop_codon 5049 CDS 12597				
pleroturus_ostreatus	fungi	171993	jgi	GTF-like	gene : CDS	JGI			1	CDS 74124 start_codon 11342 stop_codon 11492 exon 75035				
pithecopithecus_crisspa	fungi	182362	jgi	GTF-like	gene : CDS	JGI			1	CDS 76774 start_codon 13001 stop_codon 12966 exon 79621				
plutella_xylosteola	metazoa	135027	gigabid	GFF-like	mRNA : CDS	Cuff GLEAN			2	mRNA 18071 CDS 16956				
pogona_vitticeps	metazoa	231021	gigabid	GFF-like	mRNA : CDS	Cufflinks AUGUSTUS GeneWise			3	CDS 17929 UTR 31556 mRNA 19406 UTR 17030				
polysphondylium_pallidum	protists	116075	coge	GFF-like	gene : mRNA : CDS	CeCe			1	mRNA 12356 gene 12356 exon 45681 CDS 45681				
pongo_abelii	metazoa	467433	ensembl	GTF-like	gene : transcript : CDS	protein_coding MI_URNA snoRNA snRNA rRNA miRNA MI_URNA pseudogene misc_RNA			9	stop_codon 17914 CDS 207820 start_codon 18472 exon 22327				
populus_trichocarpa	plantae	540808	ensembl	GTF-like	gene : transcript : CDS	protein_coding			1	start_codon 43036 exon 23197 CDS 22231 stop_codon 43476				
posta_placenta	fungi	262763	jgi	GTF-like	gene : CDS	JGI			1	start_codon 15439 CDS 11493 stop_codon 15754 exon 11659				
pristonchus_pacificus	metazoa	585030	ensembl	GTF-like	gene : transcript : CDS	protein_coding nontranslating_CDS			2	CDS 26288 stop_codon 28457 exon 264970 start_codon 28715				
prunus_mume	plantae	175736	bjfu	GFF-like	mRNA : CDS	Cuff GLEAN GLENA			3	CDS 14436 mRNA 31399				
prunus_perseica	plantae	366941	phytotype	GFF-like	gene : mRNA : CDS	phytozomeB_0			1	mRNA 28701 three_prime_UTR 11298 gene 27864 exon 145434 live_prime_UTR 10986 CDS 142657				
psuedoschitschka_multiseries	protists	127923	jgi	GTF-like	gene : CDS	JGI			1	stop_codon 17122 exon 46821 CDS 46562 start_codon 17418				
psuedopocociles_heximelis	metazoa	167372	gigabid	GFF-like	mRNA : CDS	AUGUSTUS GeneWise exonerator:protein2genome:local genscan			1	mRNA 16998 CDS 150374				
pteropeplus_alecto	metazoa	197721	gigabid	GFF-like	mRNA : CDS	GLEAN			1	mRNA 12392 CDS 176329				
ptychocheila_flava	metazoa	328403	oistjp	GFF-like	gene : mRNA : CDS	AUGUSTUS			1	start_codon 31333 CDS 196333 gene 34638 stop_codon 31451 mRNA 34648				
puccinia_graminis	fungi	174892	ensembl	GTF-like	gene : transcript : CDS	rRNA_pseudogene snRNA RNase_MRP_RNA rRNA pseudogene			8	stop 17891 stop_codon 15981 start_codon 15924 CDS 11166				
punctularia_strigosazonata	fungi	462260	jgi	GTF-like	gene : CDS	JGI			1	start_codon 10744 stop_codon 10727 CDS 68946 exon 1843				
pusdamilia_nyererei	metazoa	976769	genbank	GFF-like	gene : mRNA : CDS	iRNAscan-SE RefSeq Gnomon			3	transcript 12496 CDS 42961 exon 461537 CDS gene segment 10 region 7236 iRNA 502 gene 24805 mRNA 36626 V_gene_segment 52 CNA match 570				
pyrgoscelis_adeliae	metazoa	502432	genbank	GFF-like	gene : mRNA : CDS	iRNAscan-SE Gnomon RefSeq			3	C_gene_segment 3 CNA match 13630 RNA 2 transcript 150 ncRNA 574 CDS 144949 gene 14440 mRNA 16914 D_cool 1_V_gene segment 4 region 31665 repeat region 2 exon 204259 RNA 153				
pyrenophora_teres	fungi	81999	jgi	GTF-like	gene : CDS	JGI			1	start_codon 11502 stop_codon 11504 exon 29546 CDS 29546				
pyripyra_yezoensis	protists	41647	nrifs	GFF-like	gene : CDS	FRA_NIRFS			1	intron 5503 CDS 13848 gene 10327 stop codon 5054 start codon 6015				
pyrus_bretschneideri	plantae	265781	gigabid	GFF-like	mRNA : CDS	GeneWise GLEAN			3	mRNA 23369 UTR 31012 CDS 20539 UTR 5_10961				
pythium_ultimo	protozoan	161176	ensembl	GFF-like	gene : transcript : CDS	repeatmasker pythiomedanabase Ensembl_Protists dust trf			5	stop 1810-10 repeat region 30607 three_prime_UTR 1366 gene 111 exon 4029 five_prime_UTR 1254				
python_molurus	metazoa	698381	genbank	GFF-like	gene : mRNA : CDS	iRNAscan-SE Gnomon RefSeq			3	mRNA 794 V_gene_segment 217 CDS 256575 transcript 323 gene 20599 CDS 25761 CNA match 1195 RNA 260 C_gene_segment 27				
rattus_norvegicus	metazoa	525298	ensembl	GTF-like	gene : transcript : CDS	pseudogene MI_URNA RNA miRNA protein_coding MI_URNA processed_pseudogene misc_RNA			8	stop_codon 22630 start_codon 22316 exon 247820 CDS 232532				
reticulomyia_filosa	protists	344883	ensembl	GFF-like	gene : transcript : CDS	RNA			1	three_prime_UTR 12 gene 39963 transcript 39963 five_prime_UTR 14 CDS 100640				
rhizopus_oryzae	fungi	150668	broadi	GTF-like	gene : transcript : CDS	RO3_FINAL_CALLGENES_3			1	stop 17404 start 17415 CDS 57683 exon 57974				
rhodotorula_graminis	fungi	103678	jgi	GTF-like	gene : CDS	JGI			1	exon 45433 start codon 6657 CDS 4861 stop codon 6727				
rhytidotherium_rufulum	fungi	89334	jgi	GTF-like	gene : CDS	JGI			1	start codon 11877 stop codon 11877 exon 38327 transcript 38327 stop codon 11877 exon 38327				
ricinus_communis	plantae	330327	phytotype	GFF-like	gene : mRNA : CDS	phytozomeB_0			1	mRNA 31221 three_prime_UTR 3904 gene 31221 CDS 13012 five_prime_UTR 4377 CDS 129291				
saccharomyces_cerevisiae	fungi	27989	ensembl	GTF-like	gene : transcript : CDS	snRNA RNA ncRNA protein_coding RNA snoRNA pseudogene			8	stop codon 17094 CDS 7050 start codon 6694				
saccoglossus_kowalewskii	metazoa	180645	metazome	GFF-like	gene : mRNA : CDS	JGI_gene_alt			1	CDS 137615 mRNA 11119 gene 5248 five_prime_UTR 1492 three_prime_UTR 11734				
saitolla_complicata	fungi	44298	jgi	GTF-like	gene : CDS	JGI			1	start codon 6677 stop codon 6559 CDS 15369 exon 15698				
salpingoeca_rosetta	metazoa	222773	broadi	GTF-like	gene : transcript : CDS	Froteropsis_gla_sp_ATCC50818_FINAL_CALLGENES_4			1	start codon 11728 stop codon 11728 exon 10300 transcript 10300 gene 23 transcript 20387 CDS 79742 RNA 23 three_prime_UTR 384 gene 30033 RNA 23 gene 24 biological_region 2706 five_prime_UTR 8352				
saproleptina_parasitica	protists	242270	ensembl	GFF-like	gene : transcript : CDS	ena ena_mrna ena_assembly_gap			9	stop codon 17094 CDS 22999 exon 23018 start codon 14214				
sarcophilus_harrisii	metazoa	484065	ensembl	GTF-like	gene : transcript : CDS	protein_coding RNA MI_URNA RNA pseudogene misc_RNA snoRNA MI_URNA RNA			1	CDS 64599 promoter 1 ncRNA 24 peptide 9401 RNA 2066 gene 10688 three_prime_UTR 4099 contig 6502 pseudogene exon 390				
schistosoma_mansoni	metazoa	117648	sanger	GFF-like	gene : mRNA : CDS	chado			1	three_prime_UTR 12 gene 39963 transcript 39963 five_prime_UTR 14 CDS 100640				
schizophyllum_comune	fungi	171411	jgi	GTF-like	gene : CDS	JGI			1	stop codon 17415 start codon 17416 CDS 57683 exon 57974				
schizosaccharomyces_pombe	fungi	32880	ensembl	GTF-like	gene : transcript : CDS	pseudogene ncRNA snRNA iRNA protein_coding snoRNA RNA			1	exon 45433 start codon 6657 CDS 4861 stop codon 6727				
scleroderma_citrinum	fungi	212487	jgi	GTF-like	gene : CDS	JGI			7	start codon 10175 exon 20699 stop codon 10210 exon 9043				
sclerotinia_sclerotiorum	fungi	77506	ensembl	GTF-like	gene : transcript : CDS	TRNA_pseudogene protein_coding tRNA iRNA snRNA snoRNA SRP_RNA			7	start codon 10175 exon 20699 stop codon 10210 exon 9043				
sebaceous_verbenaria	fungi	177848	jgi	GTF-like	gene : CDS	JGI			1	start codon 14280 stop codon 14280 CDS 7358 exon 10001				
septoria_musiva	fungi	67445	jgi	GTF-like	gene : CDS	JGI			1	start codon 17682 stop codon 17682 exon 22289 transcript 14643 five_prime_UTR 3138 CDS 122857				
serinus_canaria	metazoa	1351899	genbank	GFF-like	gene : CDS RefSeq Gnomon iRNAscan-SE RefSeq				1	gene 15348 CDS 143681 V_gene_segment 14 CDS match 10288 RNA 158 region 304400 mRNA 17728 exon 20485 transcript 46 NCRA 253 C_gene_segment 5				
serpula_lacrymans	fungi	168118	jgi	GTF-like	gene : CDS	JGI			1	CDS 71421 exon 73301 stop codon 11741 start codon 11662				
serpula_lacrymans	fungi	561797	genbank	GFF-like	gene : CDS RefSeq Gnomon iRNAscan-SE				3	cDNA match 671 exon 23790 transcript 1103 length 26125 region 16236 iRNA 1153 RNA 8 tRNA 824 CDS 242420 mRNA 33006 intron 23				
setaria_italica	plantae	524964	phytotype	GFF-like	gene : mRNA : CDS	phytozomeB_0			1	mRNA 40599 three_prime_UTR 3012 gene 3571 exon 20927 five_prime_UTR 25219 CDS 192580				
setophaga_tigrina	fungi	84489	jgi	GTF-like	gene : CDS	JGI			1	start codon 31062 stop codon 16600 CDS 160170 exon 16521				
siyambium_irio	plantae	279324	brasicabid	GFF-like	mRNA : CDS	CeCe			1	mRNA 49965 CDS 239368				
solanum_lycopersicum	plantae	416167	phytotype	GFF-like	gene : mRNA : CDS	phytozomeB_0			1	mRNA 34727 three_prime_UTR 15395 gene 34727 exon 160530 CDS 157239 five_prime_UTR 13548				
solanum_melongena	plantae	1365726	kazusa	GFF-like	gene : CDS	SME_r2_5.1_gene			1	intron 364457 gene 85446 CDS 443517 exon 472306				
solanum_tuberosum	plantae	537222	phytotype	GFF-like	gene : mRNA : CDS	phytozomeB_0			1	mRNA 51472 three_prime_UTR 35254 gene 35119 exon 190293 CDS 189881 five_prime_UTR 3502				
sorghum_bicolor	plantae	374393	ensembl	GTF-like	gene : transcript : CDS	protein_coding miRNA			2	start codon 31062 stop codon 16600 CDS 160170 exon 16521				
spathiphyllum_pallidum	fungi	26067	jgi	GTF-like	gene : CDS	JGI			1	start codon 5904 stop codon 5919 exon 7207 CDS 7037				
sphaerobolus_stellatus	fungi	389190	jgi	GTF-like	gene : CDS	JGI			1	stop codon 31742 start codon 31050 CDS 161778 exon 164620 exon 30591 start codon 15189 stop codon 14932				
sphaerophyllum_arctica	metazoa	166451	broadi	GFF-like	gene : transcript : CDS	SphA_arctica_JP510_1_V_CALLGENES_FINAL_1			1	gene 15348 CDS 143681 V_gene_segment 14 CDS match 10288 RNA 158 region 304400 mRNA 17728 exon 20485 transcript 46 NCRA 253 C_gene_segment 5				
spatophylla_pallidum	fungi	166451	oistjp	GFF-like	gene : transcript : CDS	marker			1	start codon 31062 stop codon 16600 CDS 160170 exon 16521				
staphylococcus_aureus	metazoa	1000000	genbank	GFF-like	gene : transcript : CDS	marker			1	gene 15348 CDS 143681 V_gene_segment 14 CDS match 10288 RNA 158 region 304400 mRNA 17728 exon 20485 transcript 46 NCRA 253 C_gene_segment 5				
staphylococcus_aureus	metazoa	1000000	genbank	GFF-like	gene : transcript : CDS	marker			1	start codon 31062 stop codon 16600 CDS 160170 exon 16521				
staphylococcus_aureus	metazoa	1000000	genbank	GFF-like	gene : transcript : CDS	marker			1	gene 15348 CDS 143681 V_gene_segment 14 CDS match 10288 RNA 158 region 304400 mRNA 17728 exon 20485 transcript 46 NCRA 253 C_gene_segment 5				
staphylococcus_aureus	metazoa	1000000	genbank	GFF-like	gene : transcript : CDS	marker			1	start codon 31062 stop codon 16600 CDS 160170 exon 16521				
staphylococcus_aureus	metazoa	1000000	genbank	GFF-like	gene : transcript : CDS	marker			1	gene 15348 CDS 143681 V_gene_segment 14 CDS match 10288 RNA 158 region 304400 mRNA 17728 exon 20485 transcript 46 NCRA 253 C_gene_segment 5				
staphylococcus_aureus	metazoa	1000000	genbank	GFF-like	gene : transcript : CDS	marker			1	start codon 31062 stop codon 16600 CDS 160170 exon 16521				
staphylococcus_aureus	metazoa	1000000	genbank	GFF-like	gene : transcript : CDS	marker			1	gene 15348 CDS 143681 V_gene_segment 14 CDS match 10288 RNA 158 region 304400 mRNA 17728 exon 20485 transcript 46 NCRA 253 C_gene_segment 5				
staphylococcus_aureus	metazoa	1000000	genbank	GFF-like	gene : transcript : CDS	marker			1	start codon 31062 stop codon 16600 CDS 160170 exon 16521				
staphylococcus_aureus	metazoa	1000000	genbank	GFF-like	gene : transcript : CDS	marker			1	gene 15348 CDS 143681 V_gene_segment 1				

GFF and GTF sequence features from genome projects												
<i>yarrovia_lipolytica</i>	fungi	29599	ensembl	GTF-like	gene : transcript : CDS	protein_coding tRNA snrRNA snoRNA SRP_RNA tRNA ncRNA P_RNA mRNA JGI	9	CDS:7521 exon:9128 start_codon:6502 stop_codon:6448	4	gene_id "YAL0_F00110g"; transcript_id "CAG77606"; exon_number "1"; gene_name "YAL0_F00110g"; transcript_name "YAL0_F00110g-1"; transcript_id "CAG77606"; name "grn1_1_g"; protein_id 15639; exon_number 1;		
<i>zasmidium_cellare</i>	fungi	111922	jgi	GTF-like	gene : CDS	gene	1	CDS:40101 exon:40101 start_codon:15854 stop_codon:15866	4	name "grn1_1_g"; protein_id 15639; exon_number 1;		
<i>zea_mays</i>	plantae	1149320	ensembl	GTF-like	gene : transcript : CDS	protein_coding pseudogene mRNA transposable_element	4	start_codon:111840 exon:488081 stop_codon:114530 CDS:434869	4	gene_id "GRMZM2G581216"; transcript_id "GRMZM2G581216_T01"; exon_number "1"; protein_id "GRMZM2G581216_P01";		
<i>zonotrichia_albicollis</i>	metazoa	497997	genbank	GFF-like	gene : mRNA : CDS	RefSeq Gnomon tRNAscan-SE	3	match:150 V_gene_segment:9 exon:228308 cDNA_match:2405 tRNA:159 CDS:213497 transcript:1459 C_gene_segment:1 gene:15235 mRNA:18715 region:6018	11	ID=102260483; gene_name="Znev_14974"; transcript_id="XP_002498017"; CDS:gene=LOC102260483; prod_uct=uncharacterized protein LOC102260483; protein_id=XP_002498017; gene_id=L7588_03006; gene_version="1"; transcript_id="KDR07489"; transcript_version="1"; exon_number "1"; gene_name "Znev_14974"; transcript_name "Znev_14974"; protein_id "KDR07489"; Parent=XM_002497972;1;Parent=XP_002498017;1;Name=XP_002498017;1;Alias=XP_002498017;1.ZYR0-ORF1896;ZYR0OG00132g;coge_id=41221476;codon_start=1		
<i>zootermopsis_nevadensis</i>	metazoa	199263	ensembl	GTF-like	gene : transcript : CDS	mRNA snrRNA tRNA misc_RNA RNase_MRP_RNA SRP_RNA snoRNA RNase_P_RNA protein_coding	9	start_codon:13020 exon:86393 CDS:86163 stop_codon:13687	4	gene_id "L7588_03006"; gene_version "1"; transcript_id "KDR07489"; transcript_version "1"; exon_number "1"; gene_name "Znev_14974"; transcript_name "Znev_14974"; protein_id "KDR07489"; Parent=XM_002497972;1;Parent=XP_002498017;1;Name=XP_002498017;1;Alias=XP_002498017;1.ZYR0-ORF1896;ZYR0OG00132g;coge_id=41221476;codon_start=1		
<i>zygosaccharomyces_rouxii</i>	fungi	21344	coge	GFF-like	gene : mRNA : CDS	CoGe	1	rRNA:9 exon:5541 mRNA:4991 gene:5321 ncRNA:49 gap:2 tRNA:272 CDS:5157	8	gene_id "YAL0_F00110g"; transcript_id "CAG77606"; exon_number "1"; gene_name "YAL0_F00110g"; transcript_name "YAL0_F00110g-1"; transcript_id "CAG77606"; name "grn1_1_g"; protein_id 15639; exon_number 1; gene_id "GRMZM2G581216"; transcript_id "GRMZM2G581216_T01"; exon_number "1"; protein_id "GRMZM2G581216_P01"; ID=102260483; gene_name="Znev_14974"; transcript_id="XP_002498017"; CDS:gene=LOC102260483; prod_uct=uncharacterized protein LOC102260483; protein_id=XP_002498017; gene_id=L7588_03006; gene_version="1"; transcript_id="KDR07489"; transcript_version="1"; exon_number "1"; gene_name "Znev_14974"; transcript_name "Znev_14974"; protein_id "KDR07489"; Parent=XM_002497972;1;Parent=XP_002498017;1;Name=XP_002498017;1;Alias=XP_002498017;1.ZYR0-ORF1896;ZYR0OG00132g;coge_id=41221476;codon_start=1		