

INTRONS

	Filamentous fungi <i>e.g. N. crassa</i>	S. cerevisiae	S. pombe
# of genes with introns	80%	fewer than 5%	43%
Length	Average 134 bp, mode 40-60 bp	long	81 bp average, (range 29-819), mode 48 bp
# if present	1 to many	1	1 to many
Position	anywhere, some 5' end bias	5' end	anywhere, some 5' end bias
5' splice site	GTRNGY (R = A, G) (Y = T, C)	GTRNGY (R = A, G) (Y = T, C)	GTRNGY (R = A, G) (Y = T, C)
3' splice site	YAG, Y = T, C	YAG, Y = T, C	YAG, Y = T, C
Lariat	Py GTTAACN	5' TACTAAC ^{3'}	?
Lariat to 3' sequence	14 average	31-40	?

Consensus around initiator codon.

Position	-4	-3*	-2	-1*	+1	+2	+3	+4	+5
Filamentous fungi	C	A	A/C	A/C	A	T	G	G	C
Yeast	A/Y	A	A/T	A	A	T	G	T	C
mammals	GCC	A/G	C	C	A	T	G	G	
		always a purine usually			A				

Table 2 Frequency of Bases Around the Translation Initiation Codon

Organism	Base	Position												
		-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+4	+5	+6
<i>N. crassa</i>	G	7	5	2	2	12	1	2	2	0	4	18	3	11
	A	11	11	6	12	13	8	2	30	15	18	3	6	1
	T	4	7	3	9	6	17	1	0	1	0	7	6	12
	C	9	8	20	10	2	8	29	2	18	12	4	17	7
<i>A. nidulans</i>	G	8	7	2	5	6	3	4	4	2	3	10	9	3
	A	5	9	8	8	6	8	6	26	11	15	5	5	5
	T	8	4	12	2	7	8	2	1	5	1	10	0	13
All filamentous fungi	G	8	9	7	14	10	13	20	1	14	13	6	17	7
	A	25	14	6	10	26	9	10	15	3	7	40	16	21
	T	24	37	20	27	24	19	19	76	39	50	16	18	8
	C	17	19	25	20	25	34	4	2	12	3	22	11	35
Consensus (n = 99)								C	A	C/A	A/C	G	C	

(bias)
nucleotides
around
ATG

Balance

Data were compiled from the sequences listed in Appendix 1.

The consensus sequence was assigned according to the following criteria: If the frequency of a single nucleotide is greater than or equal to 50% and greater than twice that of the second most abundant nucleotide, it is assigned as the consensus nucleotide and given in upper case. If the second criterion is satisfied but not the first, then the nucleotide is shown in lower case (see Table 3). If the sum of the frequencies is greater than 75% (but neither satisfies the above), they are jointly assigned the status of the consensus.

C > T > G > A

Balance

Table 3 Frequency of Bases at Intron Splice Sites and the Internal Site of Lariat Formation

Organism	Base	5' Splice site						Internal site of lariat formation						3' Splice site											
<i>N. crassa</i>	G	35	62	0	7	3	61	5	5	15	14	40	1	0	27	0	1	7	8	3	5	0	0	63	27
	A	4	0	0	55	28	1	3	14	21	16	17	0	0	33	63	0	32	16	20	36	2	63	0	9
	T	10	0	61	1	12	1	45	20	15	22	5	1	61	2	0	4	12	13	22	5	24	0	0	7
	C	9	1	2	0	20	0	10	19	7	11	1	61	2	1	0	58	12	21	13	12	37	0	0	15
<i>A. nidulans</i>	G	34	67	0	11	5	59	7	4	13	10	26	3	2	23	0	2	21	9	9	11	1	0	67	24
	A	4	0	0	51	25	4	6	13	15	13	24	4	0	40	67	3	15	13	13	21	4	67	0	14
	T	13	0	67	3	21	4	44	17	15	31	13	6	63	3	0	11	22	17	17	10	31	0	0	12
All filamentous fungi (n = 201)	G	97	200	0	35	16	182	18	26	46	38	97	4	7	73	0	7	40	25	20	32	2	0	201	71
	A	12	0	0	157	94	8	14	43	47	41	66	5	1	111	201	5	69	45	39	76	10	201	0	39
	T	36	0	198	6	46	10	141	52	54	90	26	20	188	9	0	29	51	49	69	27	83	0	0	28
	C	30	1	3	3	45	1	28	56	27	32	12	172	5	8	0	160	41	57	48	41	106	0	0	37
Consensus		G	G	T	A	a	G	T		t	C/A		C	T	A	A	C						C/T	A	G

Data were compiled from the sequences listed in Appendix 1.

In some cases full information was not available (see Gurr et al., 1987).

The consensus sequences were assigned according to the criteria used in Table 2.

Table 7. Codon usage in constitutively expressed genes for NADP glutamate dehydrogenase

	S. c.	N. c.	A. n.		S. c.	N. c.	A. n.		S. c.	N. c.	A. n.		S. c.	N. c.	A. n.				
Phe	UUU	5	1	2	Ser	UCU	15	3	9	Tyr	UAA	5	0	1	Cys	UGU	6	1	1
	UUC	13	17	20		UCC	14	21	23		UAC	14	15	11		UGC	0	5	2
Leu	UUA	3	0	0		UCA	2	0	0	Ter	UAA	1	1	0	Ter	UGA	0	0	1
	UUG	18	1	2		UCG	0	1	1	Ter	UAG	0	0	0	Trp	UGG	6	8	8
Leu	CUU	1	13	6	Pro	CCU	1	3	8	His	CAU	1	0	0	Arg	CGU	5	8	7
	CUC	2	22	21		CCC	0	10	10		CAC	3	9	8		CGC	1	9	11
	CUA	5	1	1		CCA	14	0	0	Gln	CAA	22	0	1		CGA	0	0	1
	CUG	0	1	2		CCG	0	0	0		CAG	0	15	14		CGG	0	0	0
Ile	AUU	11	6	6	Thr	ACU	11	8	8	Asn	AAU	5	1	1	Ser	AGU	3	0	0
	AUC	17	12	18		ACC	8	8	11		AAC	15	22	17		AGC	3	3	2
	AUA	1	0	0		ACA	2	0	2	Lys	AAA	4	0	2	Arg	AGA	14	0	0
Met	AUG	8	9	8		ACG	1	0	1		AAG	24	29	27		AGG	0	0	1
Val	GUU	15	14	16	Ala	GCU	20	15	17	Asp	GAU	4	3	5	Gly	GCU	45	31	31
	GUC	17	19	19		GCC	10	36	25		GAC	13	12	11		GGC	1	23	18
	GUA	1	0	0		GCA	2	0	1		GAA	32	0	6		GGA	1	0	3
	GUG	5	0	0		GCG	1	1	2	Glu	GAG	3	13	31		GGG	0	0	0

S. c., *Saccharomyces cerevisiae*; N. c., *Neurospora crassa*; A. n., *Aspergillus nidulans*.

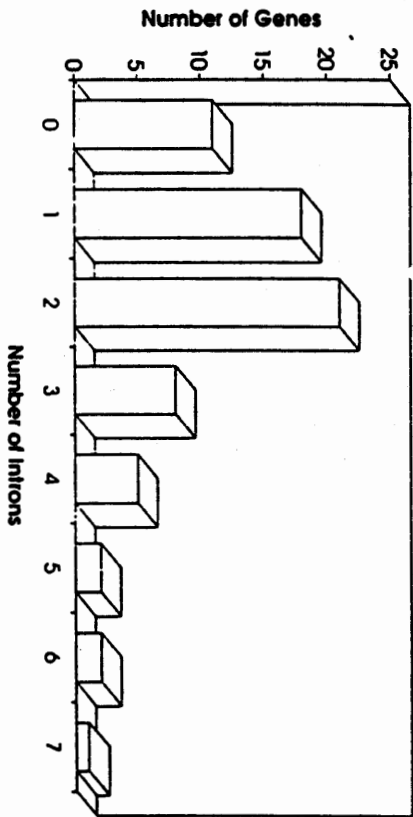


Fig. 1. Intron prevalence. The histogram shows the number of genes in the databases with 0-7 introns, as indicated on the X-axis.

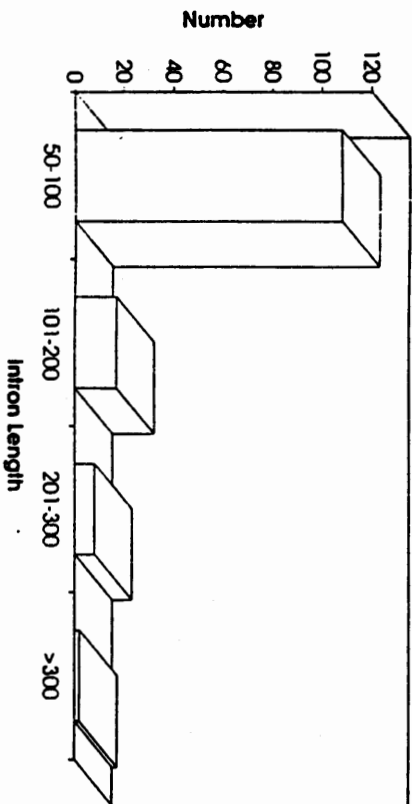
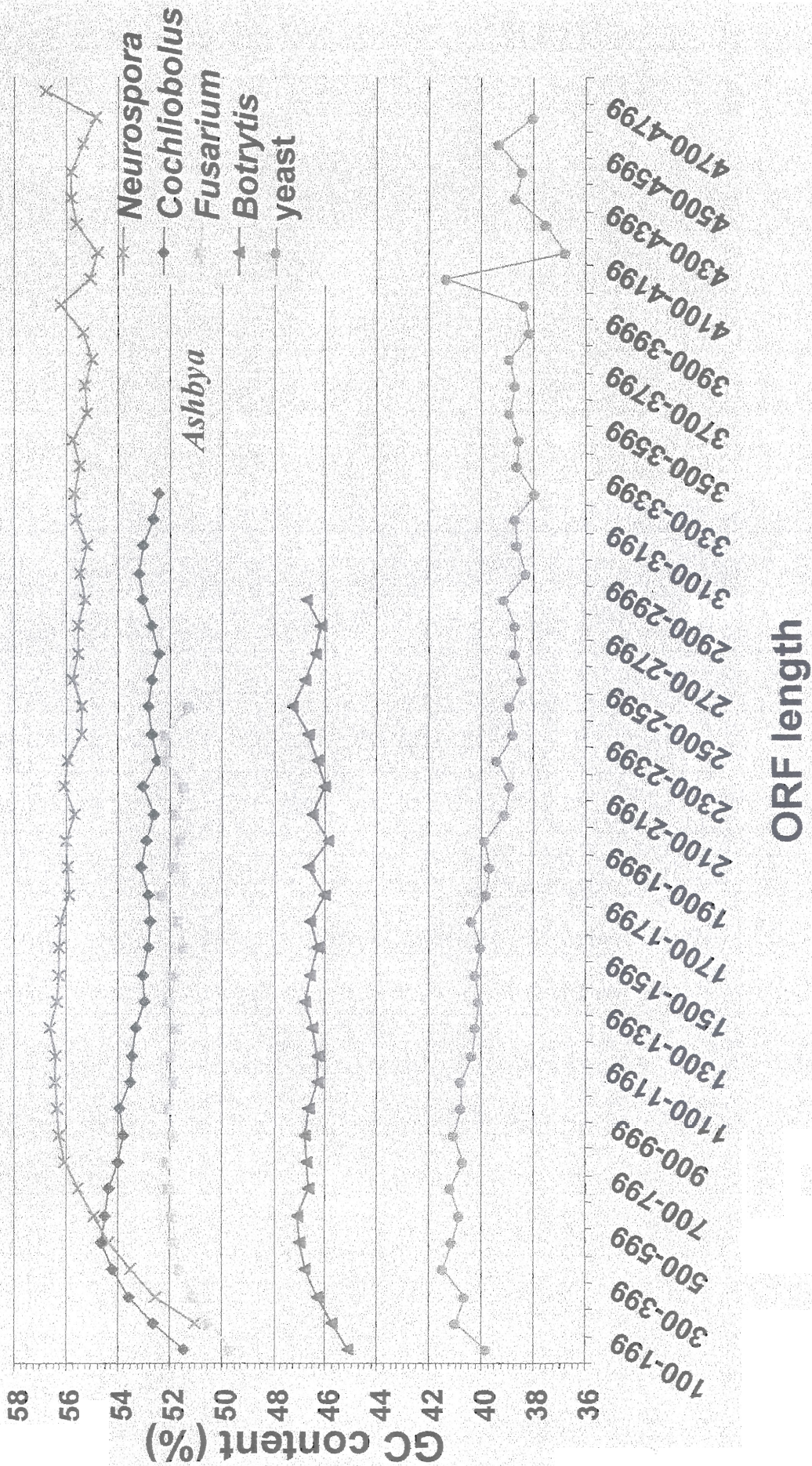


Fig. 2. Intron length. The histogram plots the number of introns of the lengths indicated on the X-axis.

*Edelmann + Staben
 Experimental Myology 18:70-81
 1994*

Percent G+C



gccounter and GC_Stat of ORFs generated by GetORFs (cutoff = 100 nucleotides)