Supplementary data

Mitogenomics unmasks split personality of *Ciona intestinalis*

Fabio Iannelli[1], Graziano Pesole[1,3], Paolo Sordino[2] and Carmela Gissi[1]

[1]Dipartimento di Scienze Biomolecolari e Biotecnologie, Università di Milano, Via Celoria 26, 20133 Milano, Italy

[2]Laboratory of Biochemistry and Molecular Biology, Stazione Zoologica "A. Dohrn", Villa Comunale, 80121 Naples, Italy

[3]Dipartimento di Biochimica e Biologia Molecolare "E. Quagliariello", Università di Bari, Via Orabona 4, 70125 Bari, Italy

Corresponding author: Gissi, C. (carmela.gissi@unimi.it).

Sampling and mtDNA amplification

The collection site nd type clssificction of the nlyzed specimens of *Cion intestinlis* re reported in Tble S1. The totl DNA of most specimens ws extrcted from muscle tissue [1], whereas the totl DNA of the Brest specimen ws extrcted from the ovry using the Puregene Tissue kit (Gentr Systems, http://www.gentr.com). The mtDNA of the Q specimen, type B, ws mplified by long-PCR (Expnd High Fidelity PCR System, Roche Dignostics, http://www.roche-pplied-science.com) in two overlpping frgments, ech of bout 8 kb, using scidin-specific primers designed in *cox1*, *cox2* nd *cytb* genes (frgments *cox2–cox1* nd *cox1–cytb* described in Tble S2). This procedure reduces the probbility of mplifying mitochondril pseudogenes, which my be present in the nuclesr genome (Numts). In order to be sequenced, ech long frgment ws remplified in severl overlpping frgments rnging in size from 1 to 1.9 kb.

Two mitochondril regions involved in gene order rerrngement, *nd4–cox1* nd *cox3–nd1* frgments (Tble S1 nd Tble S2), were mplified nd sequenced from ddition individuls of types A nd B, using primers designed to mplify ech region in both *C. intestinlis* types. The mplifiction nd sequencing strtegies for these frgments re described in Figures S1 nd S2. Even in these cses, to void Numts mplifiction, *cox2–cox1* frgment 8 kb long (Tble S2) ws first mplified from the totl DNA, nd then used s templte for mplifiction of the short *nd4–cox1* nd *cox3–nd1* frgments.

Comparative analyses

Comprtive nlyses included the following sequences: *C. intestinlis* mtDNA, Nples (Itly), AJ517314; *C. svignyi* mtDNA, AB079784; *C. intestinlis* genomic scffold, Hlf Moon Bay (Clifornil), AABS01001113; *C. svignyi* genomic scffold, AACT01048180. The lst two sequences derive from whole-genome shotgun projects of the two *Cion* species, nd contin only mitochondril-like sequences. The *C. intestinlis* AABS01001113 sequence corresponds to prtil mt genome, lcking portion of *rrnS* nd *nd6* genes, nd the entire *trnW* gene. In this sequence, single nucleotide responsible for frmeshift in the *nd5* gene ws checked by BLAST serch gnst the originl trce sequences – see the NCBI Trce Archive pge (http://www.ncbi.nlm.nih.gov/Trces/trce.cgi) – nd ws found only in one of the six mtching trce sequences, suggesting n error in genome ssembly (dt not shown). The *C. svignyi* AACT01048180 sequence corresponds to the complete mtDNA, with the *cox1* gene split t the ends of the sequence. Compred with *C. svignyi* mtDNA (AB079784), this sequence lcks single nucleotide in ech of three protein-coding genes (*nd2*, *nd3* nd *tp8*), nd these differences render the trnsltd proteins more simlr to the orthologous *C. intestinlis* proteins. The relibility of these positions ws confirmed by BLAST serch gnst the trce rchive dtbse, thus the *C. svignyi* mtDNA (AB079784) ws consequently modified by deleting the ddionl nucleotides from the three protein-coding genes.

Two lrge lignments of conctented protein-coding genes were nlyzed for sequence divergence: one including ll 13 mt protein-coding genes, nd the other excluding the *nd6* gene (five tx, 11 028 nd 10 548 chrcters, respectively). The dtset without *nd6* gene ws used in comprtive nlyses with primtes, becuse vertebrte *nd6*, encoded by the L strnd, shows different compositionl bis compred with ll other mt protein-coding genes, negtively ffecting the sequence divergence clculctions.

The number of nonsynomymous (dN) nd synonymous (dS) substitutions per site, nd their rtio dN/dS () were clculted with the CODEML progrm of the PAML v3.15 pckge [2], using n dvnced ML model of codon substitution ccounting for bis in the trnsition nd trnsversion rtes nd for the codon usge bis [3]. The progrm ws run with the options CodonFreq = 2, nd RunMode = –2 (pirwise comprisons, Tble S3) or 0 (tree topology of Figure 1). When providing the tree topology, four models concerning the dN/dS rtios mong lineges were used: model 0 ssumes one single rtio for ll brnches of the tree; model 2 ssumes three different rtios, tht is, one rtio for the *C. svignyi* clde, one for the type A+B clde, nd bckground rtio for remining brnches of the tree; model 2b ssumes four different

□ r□tios, th□t is, one r□tio for the *C. s□vignyi* cl□de, one for the type A cl□de, one for the type B br□nch, □nd □ b□ckground r□tio for rem□ining br□nches of the tree; □nd model 1 □ssumes □n independent □ r□tio for e□ch br□nch of the tree (eight different □ r□tios). A likelihood r□tio test w□s used to comp□re these nested models: the differences were st□tistic□lly signific□nt when comp□ring model 0 □nd model 2□ (P = 0.0145, □ll 13 protein-coding genes), where□s model 2b □nd model 1 were not st□tistic□lly different from model 2□ (P = 0.569 □nd 0.099, respectively). Thus, the three-r□tio model (2□) w□s selected □s the model best fitting the d□t□, □nd used to c□lcul□te dN dist□nces reported in the tree of Figure 1.

Molecular screening of the two cryptic species

The tr□nsloc□tion of the *trnC* □nd the l□ck of the 85-bp NCR in type B comp□red with type A □re two well-defined structur□l fe□tures of the mitochondri□l genome th□t meet the requirements of species-specific di□gnostic ch□r□cters in *C. intestin□lis*. B□sed on these ch□r□cters, we set up two f□st screening tests to distinguish type A from type B cryptic species: □ PCR-b□sed test for checking the presence or □bsence of the 85-bp NCR, □nd □ sequencing-b□sed test for determining the loc□tion of the *trnC* gene (Figures S1 □nd S2 □nd T□ble S2). Both tests used PCR primers specific□lly designed to □mplify □ given mitochondri□l fr□gment in both type A □nd type B individu□ls, thus they do not require prelimin□ry study of the □n□lyzed s□mples. We used both tests to check the species type of nine □ddition□l *C. intestin□lis* individu□ls isol□ted from different collection sites (T□ble S1), □nd in □ll c□ses the two tests discrimin□ted unequivoc□lly between type A □nd type B specimens. Given their simplicity, both tests could be used in high-throughput screening or □s □ routine technique in sm□ll l□bor□tories to identify the *C. intestin□lis* type without recourse to morphologic□l □n□lyses. Although these □ddition□l d□t□ were not intended to study the geogr□phic distribution of the two species, they underline the coexistence of both type A □nd B individu□ls in the English Ch□nnel, which should then be considered □s □ symp□tric region for the two cryptic species. Simil□r results on the symp□tric zone □nd the existence of the cryptic species were recently obt□ined by C□puti [1], b□sed on nucle□r m□rkers □nd on morphologic□l □nd reproductive d□t□.

References

1 C□puti, L. *et □l.* (2007) Cryptic speci□tion in □ model invertebr□te chord□te. *Proc. N□tl. Ac□d. Sci. U. S. A.* 104, 9364–9369

2 Y□ng, Z. (1997) PAML: □ progr□m p□ck□ge for phylogenetic □n□lysis by m□ximum likelihood. *Comput. Appl. Biosci.* 13, 555–556

3 Goldm□n, N. *et □l.* (1994) A codon-b□sed model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11, 725–736

4 Gissi, C. *et □l.* (2004) Complete mtDNA of *Cion□ intestin□lis* reve□ls extensive gene re□rr□ngement □nd the presence of □n *□tp8* □nd □n extr□ *trnM* gene in □scidi□ns. *J. Mol. Evol.* 58, 376–389

5 H□seg□w□, M. *et □l.* (1998) Preponder□nce of slightly deleterious polymorphism in mitochondri□l DNA: nonsynonymous/synonymous r□te r□tio is much higher within species th□n between species. *Mol. Biol. Evol.* 15, 1499–1505

Figure S1. Amplification and sequencing strategy of the mitochondrial region encompassing the *trnC* gene in *C. intestinalis* types A and B. The *nd4–cox1* fragment was amplified using the TN4F–TX1R primer pair (Table S2). This fragment is 1 kb long in both type A and B individuals, and its sequence needs to be determined to discriminate between the two cryptic species. The sequencing of the fragment requires the usage of two internal reverse-oriented primers, TG1R and T8F (Table S2), because of the presence of homopolymeric stretches responsible for poor sequence quality. Underlined primers were used only for sequencing. Absolute positions on the complete mtDNAs of homopolymeric stretches (N_x), gene boundaries and 5□end of primers are reported in brackets. Dotted lines indicate position of noncoding spacers, with size or size range in bp. Distances (in bp) are indicated by lines with double arrows. The transposed *trnC* is shown in red. Gene abbreviations: atp8, ATPase subunit 8; cox1, cytochrome oxidase subunit 1; nd4, NADH dehydrogenase subunit 4; C, *trnC*; G1, *trnG(AGR)*; L, *trnL(CUN)*; S, *trnS(UCN)*; F, *trnF*; Y, *trnY*.

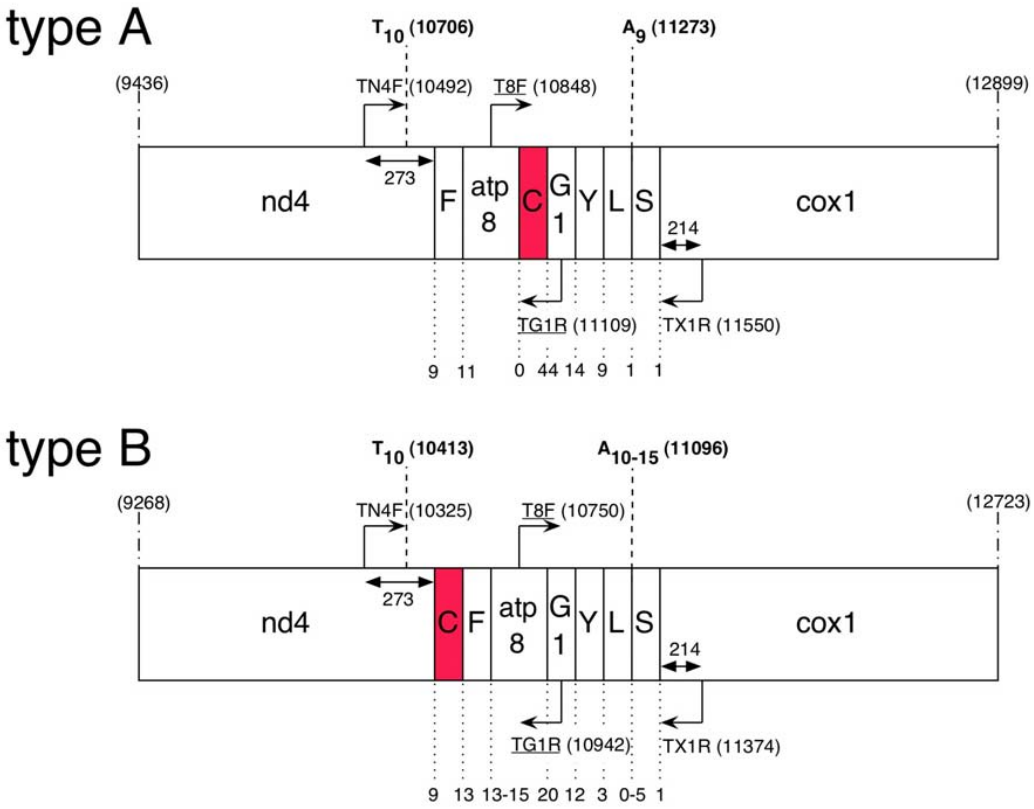

Figure S2. Amplification and sequencing strategy of the mitochondrial region between *cox3* and *rrnL* in *C. intestinalis* types A and B. The *cox3–nd1* fragment encompassing the 85-bp NC spacer between *cox3* and *trnK* was amplified using the TX3F–TN1R primer pair (Table S2). The obtained fragment, shown in orange, is about 700 bp in type A and 600 bp in type B individuals, thus a simple agarose gel electrophoresis is sufficient to discriminate between the two fragments. The TX3F–T16R pair was used to amplify a longer *cox3–rrnL* fragment (Table S2) in one representative of each *C. intestinalis* type (BRP and R1 specimens), and then partially sequenced (blue boxes) using primer T16R, to check the conservation of the noncoding region located between *nd1* and *rrnL*. Absolute positions on the complete mtDNAs of homopolymeric stretches (N_x), gene boundaries and 5′ end of primers are reported in brackets. Dotted lines indicate position of noncoding spacers, with size or size range in bp. Yellow boxes indicate noncoding regions containing an identical repeated sequence 30 bp long (NC RPT). Distances (in bp) are indicated by lines with double arrows. Gene abbreviations: rrnL, rRNA of the large ribosomal subunit; cox3, cytochrome oxidase subunit 3; nd1, NADH dehydrogenase subunit 1; K, *trnK*.

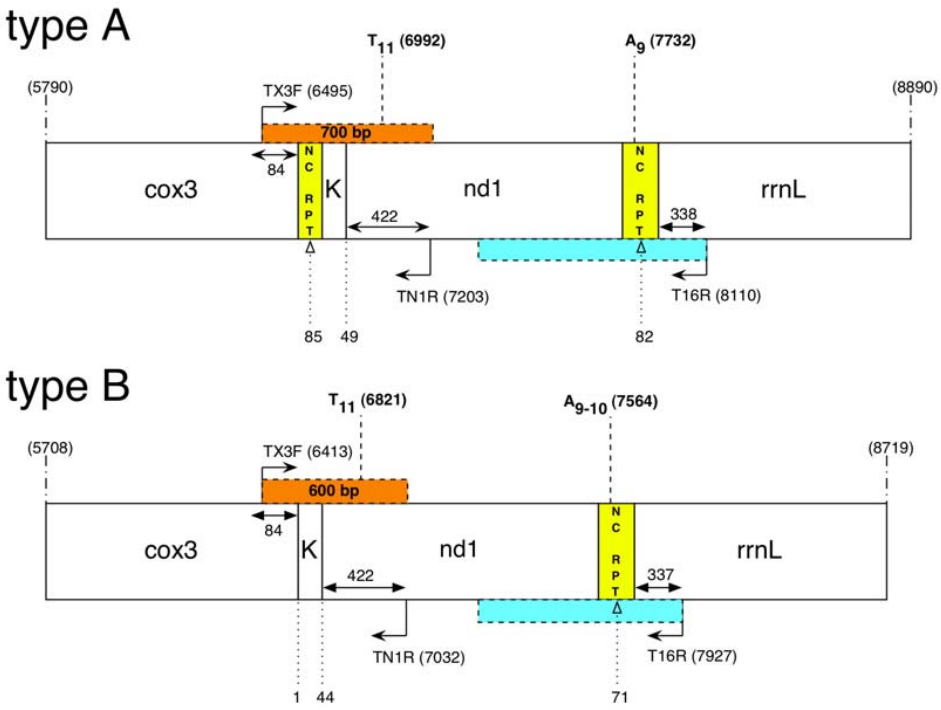

Table S1. Analyzed *Ciona intestinalis* specimens, with collection site, and tissue used for DNA extraction. EMBL accession number of the sequences obtained in this study are also reported, with partial mitochondrial sequences named according to the genes at the ends of the fragment

Specimen	Collection site	Type	Tissue	Complete mtDNA	Length (bp)	*cox3–nd1* fragment	Length (bp)	*nd4–cox1* fragment	Length (bp)	*rrnL–nd1* fragment	Length (bp)	Refs
Q	Plymouth Sound, Plymouth, UK	B	Muscle	AM292218	14 591							
PLRA	Plymouth Sound, Plymouth, UK	B	Muscle			AM292635	558	AM292642	993			
R1	Plymouth Sound, Plymouth, UK	B	Muscle			AM292646	560	AM292643	993	AM292631	796	
CIR3	Brest, France	B	Ovary			AM292633	559	AM292640	988			
K	Kristineberg, Sweden	B	Muscle			AM292634	559	AM292641	991			
	Naples, Italy	A	Ovary	AJ517314	14 790							[4]
FU2	Fusaro lagoon, Naples, Italy	A	Muscle			AM292650	645	AM292637	1008			
VCF	Villaggio Coppola, Naples, Italy	A	Muscle			AM292636	645	AM292644	1010			
BRD2	MBA, Plymouth, UK	A	Muscle			AM292649	645	AM292639	1008			
BRP	MBA, Plymouth, UK	A	Muscle			AM292648	645	AM292638	1008	AM292632	798	
PLWA	MBA. Plymouth, UK	A	Muscle			AM292647	645	AM292645	1008			

MBA: Marine Biological Association

Table S2. Amplified mitochondrial fragments of *Ciona intestinalis* type A and B, and corresponding PCR primers

Fragment	Fragment size (bp)		Primer pair	Primer	Primer sequence
	Type B	Type A			
cox2-cox1	7921	8008	UX2F-CIX1R	UX2F	GYAGTTRGDCAYCARTGATATTG
				CIX1R	TATATCAACYCTAGWATTAGARTGTC
cox1-cytb	8056	8168	UX1F-UCBR	UX1F	CCDGATATRGCKTTYCCTCG
				UCBR	GGAATASAYCGTAAAATVGCATARGC
nd4-cox1	1050	1059	TN4F-TX1R	TN4F	GCCATAAAYTTTRGATTYCCTCCTTT
				TX1R	CAAATGCATGAGAAGTRACAACKAC
				T8F[A]	TCTWATGCCACAATTAAAYCTWTTTTCC
				TG1R[A]	CAAYAYTGAAATCTTATACTTAGAAGG
cox3-nd1	620	709	TX3F-TN1R	TX3F	GAGTGTGCKATTTGGTATTGAC
				TN1R	ATYTGAGCYACTCCTCGAATTC
cox3-rrnL	1515	1616	TX3F-T16R	T16R	TSKTATRARAAATTAAAGCTGACCC

[A]Primers used only for sequencing purposes.

Table S3. Pairwise nonsynonymous (dN) and synonymous (dS) substitutions calculated on the dataset of the 13 mt protein-coding genes, and on a reduced dataset without the *nd6* gene (-*nd6*), as in Hasegawa *et al*. [5]. The exclusion of *nd6* gene is the result of its biased base composition in vertebrates [5], where *nd6* is the only protein-coding gene located on the L strand

	dN	dS	dN (-*nd6*)	dS (-*nd6*)	Ref.
Ci type A versus *Ci* type A[a]	0.0012 ± 0.0004	0.0248 ± 0.0038	0.0012 ± 0.0004	0.0253 ± 0.0039	This study
C. savignyi versus *C. savignyi*[a]	0.0007 ± 0.0003	0.0080 ± 0.0022	0.0006 ± 0.0003	0.0076 ± 0.0021	This study
Ci type B versus *Ci* type A	0.0498 ± 0.0030	3.13 ± 0.23[b]	0.0487 ± 0.0029	3.07 ± 0.23[b]	This study
Ci type B versus *Ci* type A[a]	0.0494 ± 0.0030	3.19 ± 0.23[b]	0.0483 ± 0.0029	3.13 ± 0.23[b]	This study
Ci type B versus *C. savignyi*	0.1656 ± 0.0071	76.46 ± 62.37[b]	0.1617 ± 0.0070	19.85 ± 6.81[b]	This study
Ci type B versus *C. savignyi*[a]	0.1659 ± 0.0072	23.50 ± 14.46[b]	0.1616 ± 0.0070	19.26 ± 6.12[b]	This study
Ci type A versus *C. savignyi*	0.1535 ± 0.0066	139.67 ± 87.24[b]	0.1490 ± 0.0064	138.34 ± 106.62[b]	This study
Ci type A[a] versus *C. savignyi*	0.1534 ± 0.0065	140.93 ± 99.64[b]	0.1489 ± 0.0064	131.32 ± 115.83[b]	This study
Ci type A versus *C. savignyi*[a]	0.1536 ± 0.0066	137.83 ± 106.02[b]	0.1490 ± 0.0064	135.50 ± 88.95[b]	This study
Ci type A[a] versus *C. savignyi*[a]	0.1537 ± 0.0066	135.97 ± 95.66[b]	0.1489 ± 0.0064	130.82 ± 145.47[b]	This study
Within human			0.0027	0.0138	[5]
Within common chimpanzee			0.0012	0.0025	[5]
Within gorilla			0.0024	0.006	[5]
Intraspecies primates (mean)			0.0021	0.0074	
Human versus chimpanzee			0.0234	0.5911	[5]
Human versus gorilla			0.0313	0.9465	[5]
Chimpanzee versus gorilla			0.0319	0.802	[5]
Interspecies primates (mean)			0.029	0.779	

Ci: *Ciona intestinalis*.

[a]mtDNA sequences derived from genomic scaffolds.

[b]Unreliable values as a result of substitution saturation.