Table S1. Sequence Data Used for Genome Assembly (Illumina and PacBio) and Validation (Fosmids and BACs), Related to Table 1

Illumina	**Insert size (bp)**	**Read pair (million) x 101 bp**	**Total bases (Gbp)**	**Average genome coverage (x)**
	300	42.9	8.2	10.8
	350	75.7	15.0	24.5
	400	45.5	9.0	15.2
	450	72.0	14.3	24.7
	500	149.6	28.2	38.1

PacBio	**# CLR reads**	**Total bases (Gbp)**	**mean/max read length (bp)**	**Coverage (x)**
uncorrected	2,136,706	10.9	5,093/36,663	14.5
corrected	1,369,600	7.9	6,990/33,949	12.3

fosmids	Illumina	300 fosmids (43.7 million reads, insert 300bp)
	Fosmid ends (Sanger)	285 ends from 178 fosmids (average length 719.4 bp)

BACs	Illumina	6 BACs (36.3 million read pairs, insert 300bp)
	PacBio	6 BACs (721,069 CLR reads, median length 2333 bp)
	BAC ends (Sanger)	119 ends from 61 BACs (average length 667.1 bp)

Table S2. **Properties of Gene Predictions Limited to the Germline Genome (MIC) vs. Those Lacking IESs (IES-less) or Specific to the MAC (Swart et al., 2013), Related to Figure 5**

Gene lengths are from the predicted start to stop codons and exclude UTRs.

	Feature	Number	Mean (bp)	Min (bp)	Max (bp)	% GC
MIC	Genes	810	727.8	150	3,488	34.6
	Exons	904	647.6	7	2,961	34.6
	Introns	94	40.1	29	118	23.7
IES-less	Genes	530	1,102	153	4,884	33.5
	Exons	1,147	480.7	3	4,884	34.3
	Introns	617	74.8	29	424	24
MAC (contain IESs)	Genes	16,506	1,863	150	65,451	34
	Exons	42,628	665.8	3	45,409	34.2
	Introns	26,122	90.3	28	549	23.6

Swart, E.C., Nowacki, M., Shum, J., Stiles, H., Higgins, B.P., Doak, T.G., Schotanus, K., Magrini, V.J., Minx, P., Mardis, E.R., et al. (2012). The *Oxytricha trifallax* mitochondrial genome. Genome Biol. Evol. *4*, 136–154.

Table S5. ***Oxytricha*** **MIC-Specific Protein Domains Absent from the** ***Oxytricha*** **MAC, Related to Figure 5**

Protein ID	E-value	Pfam ID	Domain name	Domain description
ctg7180000069235.g1610	1.10E-08	PF09696.4	Ctf8	Ctf8
ctg7180000090708.g2446	8.20E-06	PF02097.9	Filo_VP35	Filoviridae VP35
ctg7180000067603.g50	0.00014	PF01057.11	Parvo_NS1	Parvovirus non-structural protein NS1
ctg7180000069340.g682	5.00E-05	PF01057.11	Parvo_NS1	Parvovirus non-structural protein NS1
ctg7180000070525.g472	1.90E-05	PF01057.11	Parvo_NS1	Parvovirus non-structural protein NS1
deg7180000041353.g1621	1.70E-05	PF01057.11	Parvo_NS1	Parvovirus non-structural protein NS1
ctg7180000072061.g1205	1.10E-05	PF01057.11	Parvo_NS1	Parvovirus non-structural protein NS1
ctg7180000087953.g2027	1.00E-07	PF01057.11	Parvo_NS1	Parvovirus non-structural protein NS1
ctg7180000070723.g1844	8.30E-08	PF01057.11	Parvo_NS1	Parvovirus non-structural protein NS1
deg7180000045973.g357	6.90E-05	PF07517.8	SecA_DEAD	SecA DEAD-like domain
ctg7180000069656.g2362	5.20E-06	PF07517.8	SecA_DEAD	SecA DEAD-like domain
ctg7180000089613.g372	6.90E-07	PF07517.8	SecA_DEAD	SecA DEAD-like domain
ctg7180000067946.g928	3.40E-07	PF07517.8	SecA_DEAD	SecA DEAD-like domain

Table S7. Predicted Genes Associated with Transposable Elements, Validated by Mass Spectrometry, Related to Figure 5

Gene ID	**#Unique peptides**	**#Unique peptides with PTM**	**Modifications**	**Annotation**
deg7180000046018.g1521		1	K215(Acetyl)	*Helitron*
ctg7180000070707.g263		1	K401(Acetyl)	*Helitron*
ctg7180000076960.g1909	1			*Helitron*
ctg7180000095219.g852		2	T301(Phospho),K463(Acetyl)	*Helitron*
ctg7180000081305.g1372		1	T153(Phospho),T154(Phospho)	*Helitron*
ctg7180000070403.g1120		1	K561(Acetyl)	LINE
ctg7180000088525.g882	1	1	K172(Acetyl),K176(Acetyl)	LINE
deg7180000034613.g1354		2	T356(Phospho),T369(Phospho), K415(Acetyl),K416(Acetyl)	LINE
ctg7180000089514.g961		1	K11(Acetyl)	LINE
ctg7180000072060.g1207		1	K105(Acetyl)	LINE
ctg7180000069069.g2003		1	S269(Phospho)	LINE
ctg7180000068912.g1862		1	K538(Acetyl)	LINE
ctg7180000069518.g111		1	K112(Acetyl)	LINE
ctg7180000091356.g1584		1	K688(Acetyl)	LINE
ctg7180000069692.g1505		2	K771(Acetyl),K49(Acetyl)	LINE
ctg7180000067836.g1659	1	1	Y246(Phospho)	LINE
ctg7180000073360.g2348		1	T769(Phospho)	LINE
ctg7180000094475.g1121		1	T377(Phospho),T381(Phospho)	LINE

Table S8. Repeat Content of the MIC Genome, Related to Figure 5

	% Genome assembly coverage	**% Illumina read coverage[1]**	**% PacBio read coverage[2]**	**% Fosmid end sequence coverage[3]**
TEs				
Class 1 LINE	1.21	0.83	0.89	0
Class 2 TBE	15.6	15.0	15.2	8.43
Class 2 *Helitron*	0.33	0.26	0.33	0
Class 2 IS	0.20	0.12	0.16	0
Satellite repeats				
380bp	1.40	8.46	5.70	19.7
170bp	1.22	6.57	3.12	27.5
Others	0.65	0.69	1.52	1.12 (16.3*)
Not categorized	15.3	13.1	13.8	4.49
Total	**35.9**	**45.0**	**40.7**	**77.5**

1. 12 million randomly sampled Illumina reads (total bases 1.2Gb)
2. 145,545 randomly sampled error-corrected PacBio reads (total bases 1.1Gb)
3. 178 fosmid end sequences sequenced by Sanger method (total bases 129kb)

* Two out of 178 (1.12%) match satellite repeat families identified by RepeatModeler. Another 29 (16.3%) belong to new families of satellite repeats that RepeatModeler did not identify.