Genetic Structure of the Mating-Type Locus of *Chlamydomonas reinhardtii*

Patrick J. Ferris,[1] E. Virginia Armbrust[2] and Ursula W. Goodenough

Department of Biology, Washington University, St. Louis, Missouri 63130

Manuscript received June 15, 2001
Accepted for publication October 9, 2001

ABSTRACT

Portions of the cloned mating-type (*MT*) loci (mt^+ and mt^-) of *Chlamydomonas reinhardtii*, defined as the ~1-Mb domains of linkage group VI that are under recombinational suppression, were subjected to Northern analysis to elucidate their coding capacity. The four central rearranged segments of the loci were found to contain both housekeeping genes (expressed during several life-cycle stages) and mating-related genes, while the sequences unique to mt^+ or mt^- carried genes expressed only in the gametic or zygotic phases of the life cycle. One of these genes, *Mtd1*, is a candidate participant in gametic cell fusion; two others, *Mta1* and *Ezy2*, are candidate participants in the uniparental inheritance of chloroplast DNA. The identified housekeeping genes include *Pdk*, encoding pyruvate dehydrogenase kinase, and *GdcH*, encoding glycine decarboxylase complex subunit H. Unusual genetic configurations include three genes whose sequences overlap, one gene that has inserted into the coding region of another, several genes that have been inactivated by rearrangements in the region, and genes that have undergone tandem duplication. This report extends our original conclusion that the *MT* locus has incurred high levels of mutational change.

THE mating-type (*MT*) locus of the haploid green alga *Chlamydomonas reinhardtii*, located 30 cM from the centromere of linkage group (chromosome) VI, is involved in generating mating-type *plus* or *minus* gametic phenotypes in response to nitrogen starvation (Goodenough *et al.* 1995). The mt^+ and mt^- versions of this locus segregate 2:2 at meiosis, but early genetic analysis documented that numerous genetic markers that map to the region fail to recombine with one another, suggesting that recombinational suppression is responsible for the observed segregation patterns (Gillham 1969). This inference was confirmed with the cloning of both the mt^+ and mt^- loci (Ferris and Goodenough 1994). The locus (Figure 1) consists of an ~1-Mb region of recombinational suppression, in the center of which is an ~200-kb domain [the rearranged (R) domain] that has undergone numerous translocations and inversions involving four large segments of the domain (Figure 1). These rearrangements presumably suppress meiotic crossing over in the flanking telomere-proximal (T) and centromere-proximal (C) domains of the locus.

Of the genetic markers under recombinational suppression, three define genes that are selectively transcribed in response to nitrogen starvation and are directly involved with generating either the *plus* or the *minus* gametic phenotypes.

1. The *Fus1* gene, originally marked by the *imp1* mutation, encodes an 810-amino-acid glycoprotein that is necessary for gametic cell fusion. It is located in region *c* of the mt^+ R domain (Figure 1) and has no homolog in the mt^- locus (Ferris *et al.* 1996).
2. The *Mid* gene, originally marked by the *imp11* mutation (the mutant allele and mutant strain are henceforth designated *mid-1*), encodes a 147-amino-acid regulatory protein, related to a family of nitrogen-sensitive transcriptional regulators (Schauser *et al.* 1999), that induces cells to differentiate as *minus* gametes. It is located in region *f* of the mt^- locus and has no homolog in the mt^+ locus (Ferris and Goodenough 1997).
3. The *Sad1* gene, marked by the *imp10/imp12* mutations (Hwang *et al.* 1981) and the *agl* mutation (Matsuda *et al.* 1988), encodes a 3875-amino-acid protein that serves as the flagellar sexual agglutinin of *minus* gametes. It is located just centromere-proximal to the mt^- R domain (Figure 1), with an allele located in the homologous position in the mt^+ locus, an allele that is ordinarily not expressed because its expression is Mid-dependent and *plus* cells lack the *Mid* gene. A full report on the characterization of the *Sad1* gene is in preparation.

Although several genes involved with mating map to the *MT* locus, including several new genes that are described in this report, many other gamete-specific genes are not linked to *MT* and are designated as "autosomal" (Goodenough *et al.* 1995; Kurvari *et al.* 1998). Thus, although the *mid-1* strain carries the mt^- chromosome VI and hence lacks an mt^+ locus, it nonetheless differentiates as a *plus* gamete and requires only a *Fus1* trans-

[1]*Corresponding author:* Department of Biology, Campus Box 1229, 1 Brookings Dr., Washington University, St. Louis, MO 63130. E-mail: ferris@biosgi.wustl.edu

[2]*Present address:* Marine Molecular Biotechnology Laboratory, School of Oceanography, University of Washington, Seattle, WA 98195.

gene to achieve mating competency (Ferris and Goodenough 1997), indicating that most genes necessary for *plus* gametogenesis, including the agglutinin gene, are autosomal. To ask why some mating-related genes reside in the *MT* locus and others in autosomes under *Mid* regulation is a way of phrasing the unanswered question as to the "purpose" of the *MT* locus.

The *MT* locus is also involved in mediating uniparental transmission of organelle genomes during the zygotic phase of the *C. reinhardtii* life cycle. All four meiotic products of zygote germination ordinarily inherit chloroplast DNA (cpDNA) from the *plus* parent only and mitochondrial DNA from the *minus* parent only, the nontransmitted organellar DNAs having been selectively degraded during zygote maturation (Armbrust 1998; Remacle and Matagne 1998). The *EZY1* locus comprises seven to eight tandem iterations of a gene that is transcribed immediately after zygote formation and encodes a 414-amino-acid protein that associates with cpDNA and presumably plays some role in its selective transmission patterns (Armbrust *et al.* 1993). The *Ezy1* gene cluster is located centromere-proximal to the *Sad1* gene in both the mt^+ and mt^- loci (Figure 1).

Mutant alleles in the *MT* locus that fail to recombine also mark several genes that are expressed in vegetative (mitotic) cells and play no known specific role in gametogenesis or zygote development. Five of these "housekeeping" genes have been cloned and at least partially characterized; all lie outside the R domain at positions designated in Figure 1. The *Nic7* (nicotinamide-requiring), *Ac29* (acetate-requiring), and *Thi10* (thiamine-requiring) gene sequences have been identified by their ability to complement mutant alleles (Ferris 1995). The *Ac29* gene has been shown to encode a 495-amino-acid protein homologous to the Arabidopsis protein *ALBINO3* (Sundberg *et al.* 1997), which is involved in the biogenesis of the chloroplast light-harvesting complex (Naver *et al.* 2000), and the *Thi10* gene encodes hydroxyethylthiazole kinase (K. Shimogawara, personal communication), an enzyme in the biosynthetic pathway of thiamine. The *Mat3* gene encodes a 1209-amino-acid homolog of the retinoblastoma (Rb) protein and is involved in regulation of the cell cycle (Armbrust *et al.* 1995; Umen and Goodenough 2001b), and the *Fa1* gene encodes a 1787-amino-acid protein involved in flagellar morphogenesis (Finst *et al.* 2000).

The finding that such housekeeping genes are intermixed with life-cycle-specific genes suggests that the *MT* locus arose in an "ordinary" chromosome, in much the same way that the sex chromosomes of mammals were once ordinary chromosomes and continue to encode non-sex-related proteins (Lahn and Page 1999). However, all of the previously known housekeeping genes mapped outside the rearranged R domain, leaving open the possibility that the DNA within the R domain itself might be either restricted to sex-related functions or largely noncoding—like most of the mammalian Y chromosomes. Northern analysis of the region, reported in this article, documents that this is not the case: Genes prove to be abundant within the R domain, and many of them are expressed in vegetative cells. Therefore, the R domain of chromosome VI has been subjected to numerous local rearrangements while continuing to maintain (most of) its prior genetic activities.

We also report the characterization of several genes that are found in one *MT* locus but not the other, expanding our understanding of the coding capacity of *MT* and providing additional evidence for high mutational change in the region (Ferris *et al.* 1997).

MATERIALS AND METHODS

Northern analysis: The *C. reinhardtii* strains used to prepare RNA for Northern analysis were wild-type strains CC-620 (mt^+) and CC-621 (mt^-) all strains are available from the Chlamydomonas Genetics Center, Duke University (Durham, North Carolina). Cultures were maintained in continuous light on Tris-acetate-phosphate (TAP) medium (Harris 1989) solidified with 1.5% agar. Vegetative RNA was prepared from cells in logarithmic growth in flasks of TAP medium. Gametes were obtained by transferring cells maintained on plates for at least 7 days (Martin and Goodenough 1975) to nitrogen-free high salt minimal media (Harris 1989) for 1–2 hr. Zygotes were produced by mating equal numbers of *plus* and *minus* gametes and harvesting after 30 min or 3 hr. Preparation of Northern blots was as described (Ferris *et al.* 2001). Northern blots were stripped and reused several times during the course of these experiments. Most of the probes were prepared using restriction fragments purified from the λEMBL3 genomic phage clones that comprise the chromosome walk through the *MT* loci (Ferris and Goodenough 1994), radiolabeled with [α-^{32}P]dCTP (DuPont/New England Nuclear Research Products) by random priming.

Isolation of cDNA clones: The cDNA clones for *pr6(+)*, *pr6(−)*, *Mta1*, *Mta2*, and *Ezy2* were identified by screening plaque lifts of a cDNA expression library in Uni-ZAPXR (Stratagene, La Jolla, CA) prepared from 1-hr zygotic poly(A)+ RNA (Armbrust *et al.* 1993) by hybridization with appropriate radiolabeled genomic probes. Inserts from positive clones were excised as Bluescript SK plasmids with R408 helper phage according to the manufacturer's instructions. The cDNA that was eventually used as probe 6 was first cloned fortuitously as a consequence of its cross-hybridization to a probe derived from DNA flanking the *Mid* gene; subsequent analysis then identified the location of the corresponding gene in the T domain.

DNA sequencing and analysis: The strategy for DNA sequencing included subcloning, gene-specific primers, nested deletions using the double strand nested deletion kit (Pharmacia, Piscataway, NJ) and use of the GPS-1 genome priming system (New England Biolabs, Beverly, MA). Some sequence data were obtained by making single-stranded DNA according to Ausubel *et al.* (1989), which was used for dideoxy sequencing with the sequenase kit (United States Biochemical, Cleveland). The bulk of the sequencing was performed with the ABI PRISM dye terminator cycle sequencing ready reaction kit using double-stranded plasmid DNA and subsequent analysis on an ABI DNA sequencer. Sequence data were compiled and analyzed using the Genetics Computer Group sequence analysis software package for VAX/VMS computers (Devereux *et al.* 1984). Sequences were further investigated using

the NCBI BLAST program, the TMpred, and the COILS program (Lupas 1996).

The sequences described in this article have the following GenBank accession numbers: *Nic7* partial genomic, AY032929; *pr6(−)* cDNA, AY032930; *pr6(+)* cDNA, AY032931; *Mtd1* cDNA, AF417574; *Pr46* genomic, AF387366; *Pdk* genomic, AF387365; *Ezy2* genomic (mt^+), AF399653; Ψ-*Ezy2* genomic (mt^-), AF399654; autosomal *a* region, left border of the duplication, AF417573; autosomal *a* region, right border of the duplication (*Mta2* and *Mta3* genes), AF309495; mt^+ *a* region, left border of the duplication, AF417572; mt^+ *a* region, right border of the duplication (*Mta1*, Ψ-*Mta2*, and Ψ-*Mta3* genes), AF417571.

RNase protection analysis: Total RNA was isolated essentially as described by Kirk and Kirk (1985). Poly(A)+ RNA was isolated with the BioMag mRNA purification kit (Perseptive Diagnostics). The generation and use of the *Ezy1* antisense probe was described previously (Armbrust *et al.* 1993). The *Ezy2* probe was generated by subcloning into Bluescript II SK a 600-bp *Bam*HI/*Xho*I fragment from the coding region of the *Ezy2* cDNA. The resulting plasmid was linearized with *Sma*I, and T7 RNA polymerase was used to transcribe an antisense probe of 197 nucleotides. The protected *Ezy2* probe is 172 nucleotides. The Ambion (Austin, TX) RPA II kit was used for all RNase protection assays. Ten micrograms of total RNA was used for each RNase protection assay.

Uniparental inheritance crosses: Genetic crosses were performed using standard protocols (Harris 1989). The strains used in the control cross were CC-118 (*mt^+ sr-u-2-60*) and CC-124 (wild-type mt^-). A *mid-1 mt^- (Fus1)* cross to CC-421 (*nic7 ac29a mt^- spr-u-1-27-3*), described previously (Ferris and Goodenough 1997), generated a progeny clone (B32) *mid-1 mt^- (Fus1) spr-u-1-27-3* that was crossed to CC-1952 (wild-type mt^-). The following were added to the media as necessary: 4 μg/ml nicotinamide, 100 μg/ml spectinomycin, and 100 μg/ml streptomycin.

RESULTS

Transcriptional patterns in the *MT* locus: methodology: Northern blots containing poly(A)+ RNA from vegetative cells of both mating types, gametes of both mating types, and zygotes 30 min and 3 hr into development were prepared, and these were screened with 128 probes from the *MT* locus. The data are presented on the Genetics website at http://www.genetics.org/supplemental. The probes were chosen to give near total coverage of the R domain (~90% covered, with six gaps of 2–3 kb, and most <1 kb). The C and T domains were covered less extensively (except near the R domain borders), primarily using probes known to give single-copy bands on Southern blots (Ferris and Goodenough 1994). The T domain had 35% coverage from probe 1 to the T/R border; the C domain had 75% coverage from the R/C border to the swamp (Figure 1; *cf.* Ferris and Goodenough 1994).

From these primary data we attempted to identify all the *bona fide* genes within the R domains of the mt^+ and mt^- loci, an analysis complicated by false negatives and false positives.

False negatives (a gene failing to be identified by the Northern analysis) could result for several reasons.

1. The message is of low abundance and the blots are not sensitive enough. For example, probe 54, known to contain part of the *Mid* gene, and probe 93, known to contain part of the *Mat3* gene, did not generate Northern-blot signals under the conditions used.
2. Signals produced by cross-hybridizing repetitive sequences in the probe may obscure gene-specific signals. For example, one cannot discern the *Nic7* mRNA against the smeared hybridization signals produced by probe 5 (Figure 2A).
3. The gene may not be expressed under the growth conditions used or during the life cycle stages tested.
4. The gene may not have been represented in any of our probes, although this is unlikely for the R domain. Given these considerations, the gene density displayed in Figure 1 is very likely to be an underestimate.

False positives result if the probe cross-hybridizes to messages derived from elsewhere in the genome. This could result from repetitive-sequence elements in the probe that are present in unrelated messages, most likely in the 3′ untranslated region (UTR), or a probe that detects a transposon or a duplicated gene. In the case of duplicates, the copies in the *MT* locus might be functional (although we have no documented examples of this), or, like the genes in the *a* region of mt^+ (see below), they might be pseudogenes. In constructing Figure 1 and Table 1 we endeavored to eliminate false positives, but this is necessarily subjective. In general, multiple bands or smears were considered false positives, as were cases in which a DNA fragment known to be present in only the mt^+ or mt^- locus generated signals in RNA blots derived from both mating types.

Several regions where messages were identified by Northern blots were analyzed in more detail by DNA sequencing to confirm that the Northern analysis accurately predicts genes. This was considered particularly important to verify the existence of vegetatively expressed genes within the R domain and to identify new genes involved in the mating process. The Chlamydomonas expressed sequence tag (EST) data from the Kazusa Institute (Asamizu *et al.* 1999, 2000) and from the Chlamydomonas Genome Project (http://www.biology.duke.edu/chlamy_genome) have been very useful for the first of these purposes. Not surprisingly, however, since these libraries were derived from vegetative cells, no ESTs were identified for any of the gamete- or zygote-specific genes reported here.

Transcriptional patterns in the *mt* locus: observations: Figure 1 shows the location of major transcriptional units in the mt^+ and mt^- loci, with additional information on the various transcripts provided in Table 1. Genes designated by boxes are expressed during the vegetative phase of the life cycle; most of these were also expressed in gametes and early zygotes (Table 1). These presumably represent genes whose products function throughout the life cycle, and they are henceforth referred to as housekeeping genes. Genes designated by circles are

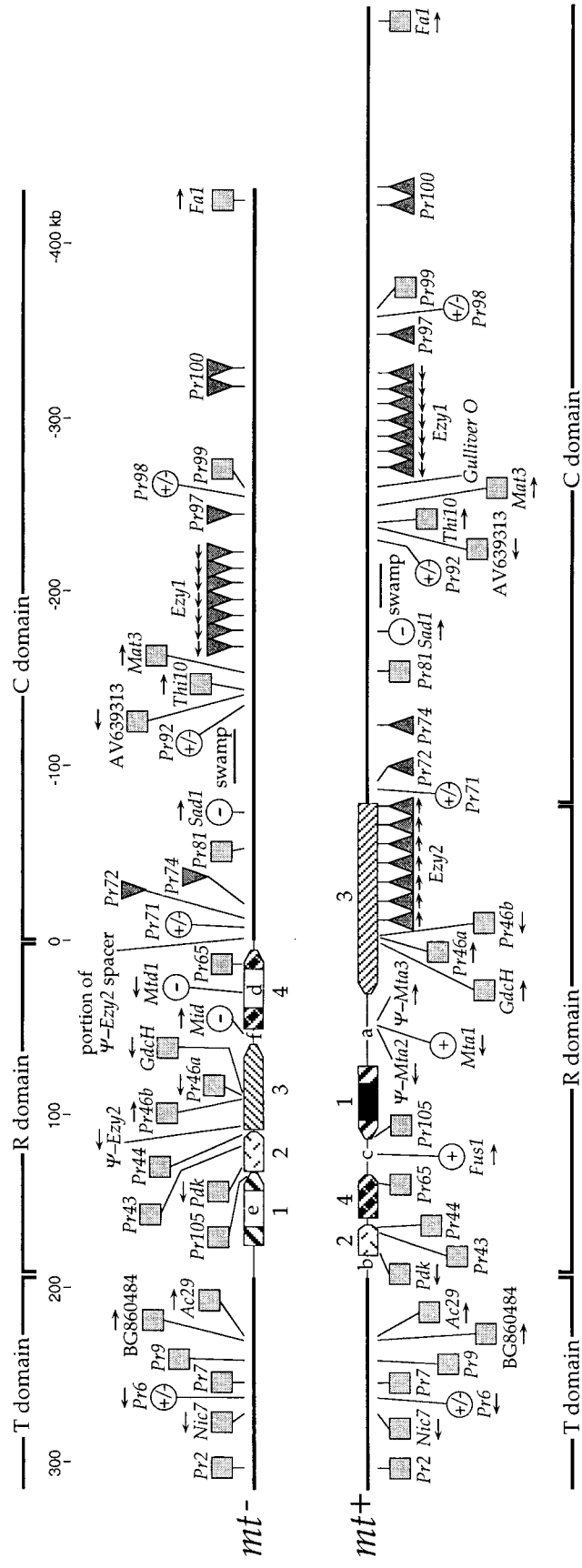

FIGURE 1.—Transcript map of the *MT* locus. The top half diagrams the *mt*⁻ locus and the bottom half the *mt*⁺ locus. The T (telomere-proximal), R (rearranged), and C (centromere-proximal) domains are indicated. The four segments of homology within the two R domains are drawn as shaded boxes, with the shape indicating their orientation. The segments are numbered 1–4. The letters a–f indicate regions within the R domain that are unique to *mt*⁻ or *mt*⁺. The genes/transcripts identified in *mt*⁻ are shown above the line, those for *mt*⁺ below the line. Squares represent messages found in vegetative cells. Circles represent messages first turned on in gametes. A + within the circle indicates a message seen only in *mt*⁺ gametes; a −, only in *mt*⁻ gametes; +/−, present in gametes of both mating types. Triangles represent messages limited to the zygote stage. Arrows indicate the direction of transcription when known (or the orientation of the gene in the case of pseudogenes). The scale (in kilobases) is indicated for *mt*⁻. Further descriptions of the genes are found in Table 1; data are given on the GENETICS website at http://www.genetics.org/supplemental.

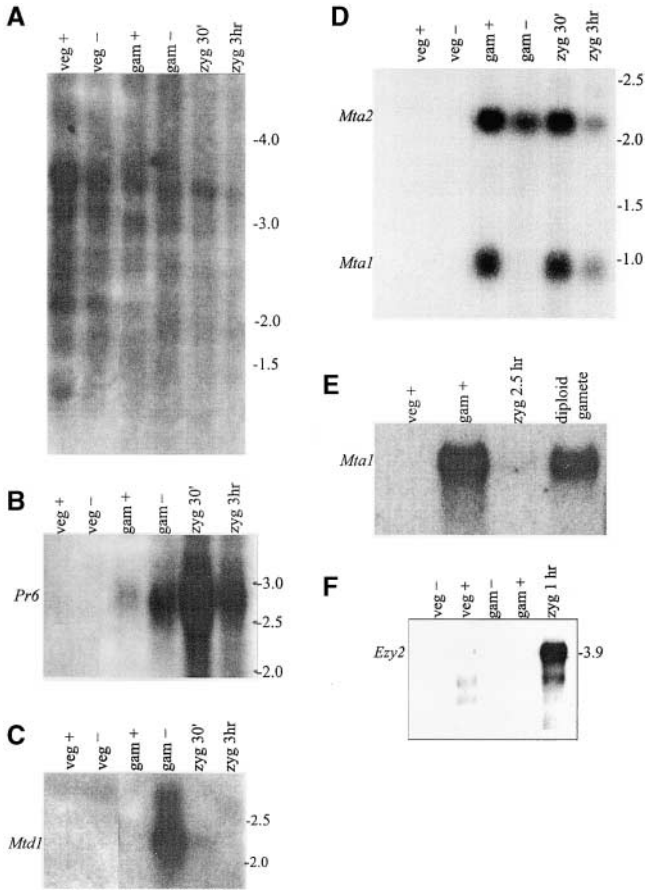

FIGURE 2.—Northern blots hybridized to selected *MT* locus probes. (A–D) Poly(A)+ RNA was isolated from mt^+ vegetative cells (veg^+), mt^- vegetative cells (veg^-), mt^+ gametes (gam^+), mt^- gametes (gam^-), zygotes 30 min after mating (zyg 30′), or zygotes 3 hr after mating (zyg 3 hr). The size of the RNA is indicated on the right (in kilobases). (A) Blot hybridized with probe 5 (*Nic7* gene). (B) Blot hybridized with probe 6. (C) Blot hybridized with probe 61 (*Mtd1* gene). (D) Blot hybridized with probes derived from the *Mta1* and *Mta2* cDNAs. (E) Blot hybridized with an *Mta1* cDNA probe. Total RNA was isolated from veg^+ and gam^+, from zygotes 2.5 hr after mating (zyg 2.5 hr), and from gametes of an mt^+/mt^- diploid (diploid gamete). (F) Poly(A)+ RNA from the designated stages hybridized with the 6.5-kb *Xho*I fragment from the 16-kb repeat of the *Ezy2* locus of mt^+ (Figure 7); the 3.9-kb *Ezy2* signal is visible only in the 1-hr zygote sample (the minor band beneath it is assumed to be artifactual since it is not always present; *cf.* Figure 9).

expressed only in gametes, with (+) transcripts found only in *plus* gametes, (−) transcripts found only in *minus* gametes, and (+/−) found in both; the expression of these genes is presumably regulated directly or indirectly by nitrogen starvation. Genes designated by triangles are expressed only in early zygotes (whether their expression continues into the late stages of zygote development/germination has not been investigated); the expression of these genes is presumably regulated directly or indirectly by gametic cell fusion (MINAMI and GOODENOUGH 1978; FERRIS and GOODENOUGH 1987).

The following sections describe genes or *MT* regions that were subjected to in-depth analysis.

***Nic7*:** Probe (Pr) 5 is a 2.1-kb genomic fragment from the center of the *Nic7* gene in the T domain, as defined by rescue of the *nic7* mutation using transformation (FERRIS 1995). Probe 5 hybridizes to several bands in Northern blots (Figure 2A), precluding identification of the *Nic7* transcript. The 2079-bp DNA sequence of this probe was determined (GenBank no. AY032929). A segment of the $(GT)_n$ repeat (KANG and FAWLEY 1997) in the fragment may be responsible for the cross-hybridization. (GT repeats are commonly encountered in the EST libraries.)

Database searches with this partial *Nic7* sequence yielded no matches to *C. reinhardtii* ESTs (which is not surprising given that *Nic7* is probably a low-abundance message, and the sequence is not near the 5′ or 3′ ends). After excluding six putative introns from the Chlamydomonas sequence, a significant homology (63% identity) was found to an Arabidopsis protein predicted from genomic sequencing (GenBank no. BAB09392). The function of the Arabidopsis gene is unknown. However, both sequences display a weak homology to prokaryotic quinolinate synthetase A genes (*e.g.*, 24% identity to the *Escherichia coli nadA* sequence). Since quinolinate synthetase participates in one pathway of NAD biosynthesis (MAGNI *et al.* 1999), and since *nic7* mutants require nicotinamide, the *Nic7* gene (and its Arabidopsis counterpart) may code for this enzyme.

***Pr6*:** The *Pr6* gene, detected by a cDNA called probe 6, is not expressed in vegetative cells but is transcribed at low levels in gametes, abundantly in 30-min zygotes, and somewhat less abundantly in 3-hr zygotes (Figure 2B). The gene is located in the T domain (Figure 1).

Two distinct classes of *Pr6* cDNA clones, with slightly different sequences, are present in 1-hr-zygote cDNA libraries, indicating that both the mt^+ (CC-620 parent) and mt^- (CC-621 parent) alleles are expressed. Restriction-site polymorphisms allowed us to assign the two cDNA types to their respective alleles (FERRIS and GOODENOUGH 1994), hereafter called *pr6(+)* and *pr6(−)*. The *pr6(−)* cDNA encodes a 721-amino-acid protein (GenBank no. AY032930), whereas the *pr6(+)* cDNA (GenBank no. AY032931) contains an extra 8 bp in its coding region, generating a stop codon-producing frameshift that would result in a truncated 455-amino-acid protein. Synonymous and nonsynonymous codon differences also differentiate the two alleles (Table 2).

A single recombinant between *nic7* and *ac29* has been isolated (SMYTH *et al.* 1975), and this *nic7 ac29a* mt^- strain (CC-350) and its derivatives (including CC-421) contain the *pr6(+)* allele in an mt^- strain (confirmed by PCR amplification and sequencing). Zygotes produced in crosses between CC-350 derivatives and mt^+ strains are viable despite the fact that they carry two copies of the *Pr6(+)* frameshift allele and no copies of the *Pr6(−)* allele, indicating that the Pr6 protein either is nonessential for zygote maturation (at least in the laboratory) or

TABLE 1

Genes identified in the mating-type loci

Gene name	Message size (kb)	When expressed	*B* value[a]	% GC[b]	Comments
			Genes in both mating types, T domain		
Pr2	1.9	All stages	—	—	
Nic7	ND	ND, presumed vegetative	0.61[c]	69[c]	Nicotinamide-requiring mutant; possible quinolinate synthetase A
Pr6	2.2	Low levels in gametes, higher in zygotes	0.35[d]	68[d]	
Pr7	2.2	All stages	—	—	
Pr9	2.3	All stages	—	—	
BG860484	ND	Vegetative; ND for gametes, zygotes	—	—	ESTs BG860484, BG859735, and BG856733 match the *AC29* genomic sequence adjacent to the predicted *Ac29* gene
Ac29	2.5	Vegetative; reduced or absent in gametes, zygotes	0.36	67	Acetate-requiring mutant, yellow color. Similar to the Arabidopsis *ALBINO3* gene
			Genes in both mating types, R domain		
Pr105	1.5	Vegetative; reduced or absent in gametes, zygotes	—	—	Located in segment 1
Pdk	2.6	Vegetative; no data on gametes or zygotes	0.38	65	Similar to pyruvate dehydrogenase kinase (or to branched-chain α-keto acid dehydrogenase kinase). Located in segment 2
Pr43	1.5	All stages	—	—	Located in segment 2
Pr44	2.0	All stages	—	—	Located in segment 2
Pr46a	1.1	Vegetative; reduced or absent in gametes, zygotes	0.30	56	Similar to a widely conserved protein of unknown function. Located in segment 3
Pr46b	1.4	ND, presumed vegetative	0.35	65	Located in segment 3
GdcH	1.1	Vegetative; reduced or absent in gametes, zygotes	0.55	63	Similar to glycine decarboxylase complex subunit H. Located in segment 3
Pr65	1.1	All stages	—	—	Located in segment 4
			Genes in both mating types, C domain		
Portion of Ψ-*Ezy2* spacer	NA	NA	—	—	~1.5 kb of the 16-kb repeat containing the *Ezy2* gene is at the start of the C domain region (see Figure 10)
Pr71	3.1	Gametes only	—	—	
Pr72	1.0	Zygotes only	—	—	May be a duplicate copy of the *Pr74* gene
Pr74	1.0	Zygotes only	—	—	May be a duplicate copy of the *Pr72* gene
Pr81	1.1	Vegetative; in lesser amounts in gametes, zygotes	—	—	
Sad1	12	*minus* gametes only, some remain in 30-min zygotes	—	—	*minus* agglutinin gene
Swamp	NA	NA	—	—	Section of multiple tandem repeats of 1.1-kb unit
Pr92	3.0	Gametes only	—	—	
AV639313	ND	Vegetative	—	—	ESTs AV639313 and AV634214 match the *THI10* genomic sequence adjacent to the predicted *Thi10* gene

(*Continued*)

TABLE 1

(Continued)

Gene name	Message size (kb)	When expressed	*B* value[a]	% GC[b]	Comments
Thi10	ND	ND, presumed vegetative	0.47	70	Hydroxyethylthiazole kinase gene, predicted from genomic sequence (K. SHIMOGAWARA, personal communication)
Mat3	4.5	All stages (J. G. UMEN, personal communication)	0.38	68	Retinoblastoma protein homolog
Ezy1	2.2	Zygotes only	0.44[d]	71	Zygote-specific protein that localizes to chloroplast nucleoids; multigene family
Pr97	2.2	Zygotes only	—	—	
Pr98	3.0	Gametes and zygotes	—	—	
Pr99	1.3	Vegetative; in lesser amounts in gametes, zygotes	—	—	
Pr100	6.5	Zygotes only	—	—	Probe 100 is derived from an inverted tandem duplication
Fa1	5.8	Vegetative; no data on gametes or zygotes	0.33	75	Flagellar autotomy gene
			Genes only in mating type *minus*		
Ψ-*Ezy2*	NA	NA	—	—	A single pseudogene copy of *Ezy2*
Mid	1.1	*minus* gametes	0.16	50	*minus* dominance gene
Mtd1	2.2	*minus* gametes and faintly in 30-min zygotes	0.30	68	
			Genes only in mating type *plus*		
Fus1	3.0	*plus* gametes	0.05	48	*plus*-specific fusion gene
Mta1	0.8	*plus* gametes and 30-min zygotes, less in 3-hr zygotes	0.46	65	
Ψ-*Mta2*	NA	NA	—	—	Pseudogene copy of the HRGP *Mta2* gene
Ψ-*Mta3*	NA	NA	—	—	Pseudogene copy of the *Mta3* gene
TOC2	NA	NA	—	—	A TOC2-like element (shown in Figure 5 only)
Gulliver	NA	NA	—	—	A Gulliver-like element (shown in Figure 5 only)
Ezy2	3.9	Zygotes only	0.35	68	Tandemly repeated gene family
Gulliver O	ND	ND	—	—	A copy of the Gulliver transposon, present at this position in strain CC-620 only

ND, not determined; NA, not applicable.

[a] *B* value is a measure of codon usage bias (LONG and GILLESPIE 1991). A *B* value of 0 indicates that all codons are used equally frequently; a value of 1 indicates that only one codon is used for each amino acid.

[b] Calculated for the coding region only.

[c] Based on partial gene sequence.

[d] Calculated for *pr6(−)* allele.

[e] Calculated for L20946.

TABLE 2

Level of homology between gene pairs

Section compared	Length (bp) (first gene/second gene)	Indels	% change	Codons[a] (S/N)
	Probe 6 gene *pr6(−) vs. pr6(+)*			
Coding[b]	2166/2183	3	1.6	18/15[c]
3′ UTR	528/438	4	5.6	NA
	Mta2 gene (autosomal gene *vs.* mt^+ pseudogene)			
5′ UTR	179/180	1	5.0	NA
Introns	646/656	5	3.4	NA
Coding	1032/1025	1	2.0	10/9
3′ UTR	659/720	1	3.3	NA
	Mta3 gene (autosomal gene *vs.* mt^+ pseudogene)			
5′ UTR	204/204	0	3.4	NA
Intron	360/330	4	4.7	NA
Coding	486/486	0	1.0	3/2
3′ UTR	1114/1129	6	4.6	NA
	Ezy2 gene (mt^+ gene *vs.* mt^- pseudogene)			
5′ UTR	341/322	3	3.1	NA
Introns	2484[d]/1978	29	5.2	NA
Coding	2436/2439	14	3.8	15/66
3′ UTR (part)	303/298	10	5.4	NA
	YptC4 gene (137c *vs.* S1C5[e])			
Intron 6	485/480	2	3.3	NA
	Gp1 gene (CC-621 *vs.* S1C5[f])			
Introns	300/305	2	3.7	NA
Coding	488/488	0	1.6	4/4

[a] Number of codons that contain base pair changes, shown as a pair of numbers—the first is the number of changed codons that are synonymous (S); the second is the number of nonsynonymous (N) codons.
[b] The coding portion is defined as the longer coding region of the *pr6(−)* allele.
[c] 12 synonymous changes before the frameshift, 6 after; 6 nonsynonymous changes before the frameshift, 9 after.
[d] The internal duplication in the mt^+ gene was scored as a single indel within an intron.
[e] GenBank U13167 and U55893.
[f] Our unpublished data.

remains functional in its truncated form. Uniparental inheritance of chloroplast markers occurs normally in crosses using the *Pr6(+)*-carrying strains (our unpublished results).

The sequence of Pr6p is 40% identical, over 190 amino acids, to *E. coli* Endopeptidase IV (P08395), the signal peptide peptidase (Ichihara *et al.* 1986); a comparable level of similarity is found to an Arabidopsis EST (GenBank no. AAF24059). The homology resides in the C-terminal portion of the Pr6p protein that is presumed to be missing from the *Pr6(+)* frameshifted version.

Pyruvate dehydrogenase kinase: The results using probes 37–39 highlight the problems of false positives and negatives. Probe 38 hybridizes to a 1.6- and a 1.9-kb message; probe 39 hybridizes to a 3.2- and a 1.1-kb message; probe 37, which partly overlaps probe 38, gives a negative result. The four signals all appear to be false positives; in fact, the message for the protein encoded in this region is not visualized.

We sequenced a 5813-bp region from segment 2 of the mt^- R domain (Figure 1) that covers the region represented by these three probes plus a few hundred flanking nucleotides. A BLAST search identified three *C. reinhardtii* ESTs to this sequence—two from the Chlamydomonas Genome Project mt^+ set (AW758420 and AW758419 are the 5′ and 3′ ends, respectively, of the same clone) and one from the Kazusa mt^- set (AV643090)—which all correspond to the same mRNA. We sequenced the AV643090 clone completely to identify intron borders and the 3′ end of the gene. This region contains a 4974-bp gene (GenBank no. AF387365) predicted to produce a message of ~2.6 kb, which does not correspond to any of the bands seen on Northerns.

The predicted protein product is homologous to both pyruvate dehydrogenase kinase (Pdk) and the closely related branched chain α-keto acid dehydrogenase kinase (Bckdk), containing all the conserved motifs (Thelen *et al.* 1998). Since Pdk has been characterized in plant mitochondria whereas Bckdk has not yet been identified in plants, we have opted to call the gene *Pdk*.

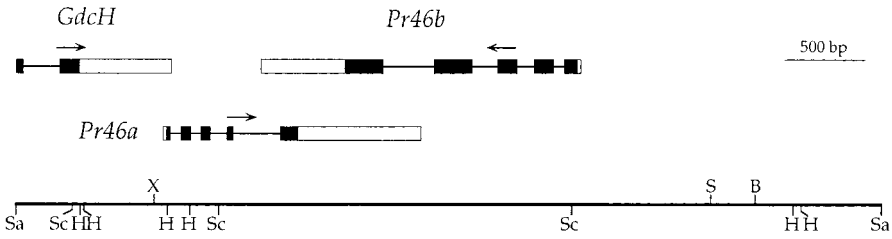

FIGURE 3.—Genomic structure of the *Pr46/GdcH* region of segment 3. Open boxes represent untranslated regions; solid boxes represent coding sequences; thin lines represent introns. Only the 3′ half of the *GdcH* gene is within the sequenced region. Arrows indicate direction of transcription. Key to the restriction sites used here and in Figures 5 and 7: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sma*I; Sa, *Sal*I; Sc, *Sac*I; X, *Xho*I; and Xb, *Xba*I.

The *Pdk* gene resides near one end of segment 2 (Figure 1) such that its 3′ UTR extends beyond the sequence discontinuity that marks the end of segment 2. This means that the final 146 bp of the mt^- 3′ UTR and the final 125 bp of the mt^+ 3′ UTR are unrelated.

Glycine decarboxylase complex subunit H, *Pr46a*, and *Pr46b*: Probe 46, a 5.4-kb Sal fragment from segment 3 of the mt^+ R domain, hybridizes to a single 1.1-kb transcript seen only in vegetative cells; however, the sequence of probe 46 matched ESTs representing three different transcripts, two of 1.1 kb and one of 1.4 kb (Figure 3). In mt^+ and mt^- genomic Southern blots, probe 46 hybridized to a single band, indicating that these three genes are present only in the *MT* locus.

The sequence of the leftmost message (Figure 3) encodes the mitochondrial enzyme glycine decarboxylase complex subunit H (GdcH), which participates in photorespiration (OLIVER 1994). The 3′ half of the *GdcH* gene is included in probe 46, and it is well represented in the *C. reinhardtii* EST database. Several alternative poly(A) addition sites are represented in the EST collections, the most common (shown in Figure 3) located within the first intron of the adjacent gene (*Pr46a*). That is, there is partial overlap between the *GdcH* and *Pr46a* transcripts (Figure 3).

Gene *Pr46a* is represented by four ESTs, all from the Kazusa collection and hence derived from the mt^- allele of the gene. One of these (AV390703) was sequenced to determine the intron locations and the 3′ end. The predicted Pr46a protein of 96 amino acids is highly conserved (80% identity to an Arabidopsis protein, 75% identity to a *Caenorhabditis elegans* protein) but of unknown function. The sequence does not appear in the yeast genome. A number of polymorphisms exist between the mt^+ and mt^- alleles, only one of which is in the coding region, resulting in an Ile in mt^+ and a Thr in mt^- at position 69, a poorly conserved region of the protein.

Gene *Pr46b* is represented by a single EST in the Kazusa collection (AV626473); this was sequenced to determine the positions of the two introns and the 3′ end. One EST from the Chlamydomonas Genome Project collection confirmed the 3′ end, and a second includes additional 5′ sequence. The predicted Pr46b protein of 267 amino acids shows 30% identity to a human cDNA (GenBank no. AK023156) and its mouse homolog (GenBank no. AK006639), of unidentified function. Again there are polymorphisms between the mt^+ genomic sequence and the mt^- cDNA: The six changes in the sequenced coding regions are all synonymous, suggesting that the gene is under selection.

Remarkably, the *Pr46a* and *Pr46b* mRNAs also overlap, in this case by 1005 bp: The 3′ end of one message is within the last intron of the other gene and vice versa (Figure 3). The 3′ UTR of each message overlaps part of the 3′ UTR and part of the coding region of the other, but there is no overlap of their coding regions.

Region *f*: In a previous publication we documented that the *Mid* gene, marked by the *mid-1* mutation, resides in region *f*, which is flanked by segments 3 and 4 and unique to the mt^- R domain (FERRIS and GOODENOUGH 1997; Figure 1). Subsequently, Christoph Beck and colleagues generated a strain (CC-3712) with a deletion that covers all of region *f* plus 8–9 kb of segment 3 and 10–12 kb of segment 4 (our unpublished data). The deletion mutant (*mid-2*) has the expected pseudo-*plus* sterile phenotype of a *mid* mutant (GOODENOUGH *et al.* 1982) but undergoes apparently normal vegetative growth under laboratory conditions. Since no transcripts other than *Mid* hybridize to the regions deleted in *mid-2*, these regions, corresponding to probes 51–59 and 103, may be free of other genes.

Gene *Mtd1*: Region *d* is a single-copy sequence found only within segment 4 of the mt^- locus (FERRIS and GOODENOUGH 1994; Figure 1). Probe 61, a restriction fragment from region *d*, hybridizes to a 2.2-kb mRNA, found in *minus* but not *plus* gametes and barely visible in the 30-min zygote sample (Figure 2C). The cognate gene for this message is called *Mtd1*. Several *Mtd1* cDNA clones were isolated, one of which was sequenced. The 2274-bp message codes for the 625-amino-acid protein shown in Figure 4. No homologs of this protein are in current databases.

Whatever function the Mtd1 protein provides to mt^- gametes, it cannot be essential in the laboratory, since mt^+ gametes transformed with the *Mid* gene can mate as *minus* and produce meiotic progeny with a mt^+ partner even though the *Mtd1* gene is absent from both parents (FERRIS and GOODENOUGH 1997). Fusion between the gametes in such crosses is very slow, however, which suggests a role for Mtd1 in efficient cell fusion, perhaps as a component of the membrane overlying the mt^- mating structure (WEISS *et al.* 1977; GOODENOUGH *et al.* 1982).

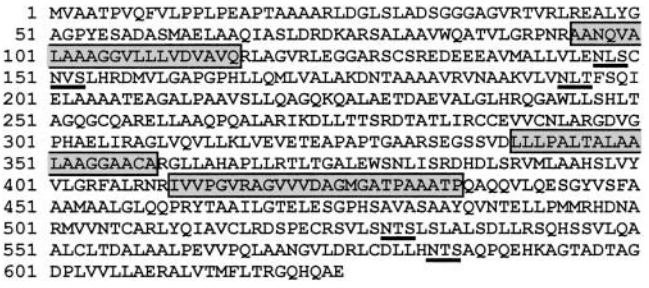

FIGURE 4.—Predicted sequence of the Mtd1 protein. Predicted transmembrane domains are in shaded boxes; canonical N-glycosylation sites are underlined.

The predicted Mtd1 protein (Figure 4) has five NXT/S N-glycosylation consensus motifs and three predicted transmembrane segments, which, if threaded sequentially, would place the NXTs in an exterior orientation.

Genes *Mta1*, *Mta2*, *Mta3*, Ψ-*Mta2*, and Ψ-*Mta3*: Region *a* was originally defined (FERRIS and GOODENOUGH 1994) as a 20-kb sequence between segments 1 and 3 in the mt^+ R domain that is not present in the mt^- R domain. However, a homologous sequence is present in an autosome, meaning that *plus* cells carry two copies of the sequence and *minus* cells carry one. The extent of the duplicated sequence has been defined by comparing restriction maps, cross-hybridizing probes, and sequencing selected sections (Figure 5). Sequences that join the duplicated mt^+ region to segments 1 and 3 (Figure 5) are absent from both the mt^- locus and from the autosomal domain.

Eleven probes (111–121 in Figure 5) were used in the Northern analysis. Of those flanking the duplicated *a* region, probes 111 and 112 detect the same 6.5-kb RNA, probe 119 detects a 3.0-kb RNA, and probe 120 detects a 0.9-kb RNA. However, since these signals are present in mt^- lanes as well, we interpret them to be false positives.

Probe 118, which lies within the duplicated *a* region, detects a 1.8-kb message at all life-cycle stages analyzed. The entire probe has been sequenced, and guided by EST matches we found that this message derives from a gene we call *Mta3*. The gene (GenBank no. AF309495) has one intron and encodes a predicted gene product of 166 amino acids, with a molecular weight of 18.5 kD, an isoelectric point (p*I*) of 11.3, and no homologs in the database. Since the *Mta3* sequence lies within the duplicated region, the *Mta3* ESTs from the vegetative *plus* library could have originated either from the autosomal copy or from the copy in the mt^+ locus. However, the three ESTs analyzed all contain sequence polymorphisms specific to the autosomal copy, suggesting that the mt^+ copy may not be transcribed. This inference is supported by the finding that the mt^+ copy carries a mutation that deletes the intron 5′ splice site so that an alternative splice junction would have to be used for the mt^+ *Mta3* gene to be functional. Our working assumption, therefore, is that the mt^+ copy of *Mta3* is a pseudogene, Ψ-*Mta3*, and that the expressed *Mta3* gene is autosomal.

Probe 117, which also lies within the duplicated *a* region, detects two messages—one 2.2 kb and one 0.8 kb—both of which are absent from vegetative cells, present in gametes and 30-min zygotes, and at reduced levels in 3-hr zygotes (Figure 2D). The 2.2-kb species is present in gametes of both mating types, whereas the 0.8-kb species is present in *plus* gametes only (Figure 2D). cDNA clones that correspond to each have been isolated.

The smaller 0.8-kb message derives from a gene we call *Mta1*, which is present in the mt^+ copy of the *a* region but absent from the autosomal copy. The *Mta1* gene is expressed in mt^+/mt^- diploid gametes (Figure 2E), indicating that its expression is not repressed by the Mid protein (diploids differentiate as *minus* gametes; EBERSOLD 1967). The mt^+-unique gene *Fus1* is also expressed in diploid gametes (FERRIS *et al.* 1996), suggesting that gamete-specific genes unique to the mt^+ locus have lost, or never acquired, Mid repressibility.

The *Mta1* gene encodes a predicted 126-amino-acid protein, Mta1 (Figure 6), of 14.6 kD, p*I* 7. Its C terminus is predicted to adopt a coiled-coil motif, generating BLAST matches to proteins such as lamin B. Amino acids 48–102, the main components of the coiled-coil domain, comprise five imperfect repeats of an 11-amino-acid sequence (Figure 6). A strikingly similar 11-amino-acid repeat domain is found in the ROPE protein of *Plasmodium chabaudi* (WERNER *et al.* 1998), where the motif is proposed to form a leucine histidine-zipper that interacts with other proteins.

The larger message detected by probe 117 derives from a gene that is expressed from the autosomal *a* region in gametes of both mating types but not vegetative cells. As reported elsewhere (FERRIS *et al.* 2001), this gene encodes a 386-amino-acid hydroxyproline-rich glycoprotein of unknown function. In our previous publication (FERRIS *et al.* 2001) we called the gene *a2* and the protein A2; using the nomenclature adopted for the present article, we call the gene *Mta2* and the protein Mta2.

When the autosomal and mt^+ genomic sequences are compared, it becomes clear that the *Mta1* coding region has been inserted into the *Mta2* gene in the mt^+ locus (Figure 5): The promoter region, the 5′ UTR (and its intron), and the first nine codons of *Mta1* correspond to the *Mta2* sequences in the autosome, after which the two sequences diverge completely, with the rest of the *Mta1* sequence being totally unrelated to the autosomal *Mta2* sequence. Downstream of the 3′ end of the *Mta1* gene, *Mta2* sequences pick up again: Although most of the second *Mta2* exon and a portion of its second intron are missing, the remainder of the gene is present. Since these *Mta2* sequences are not included in the *Mta1* transcript, this means that transcriptional termination signals downstream of the *Mta1* gene prevent expression of the adjacent *Mta2* sequences. We therefore designate this region as Ψ-*Mta2*.

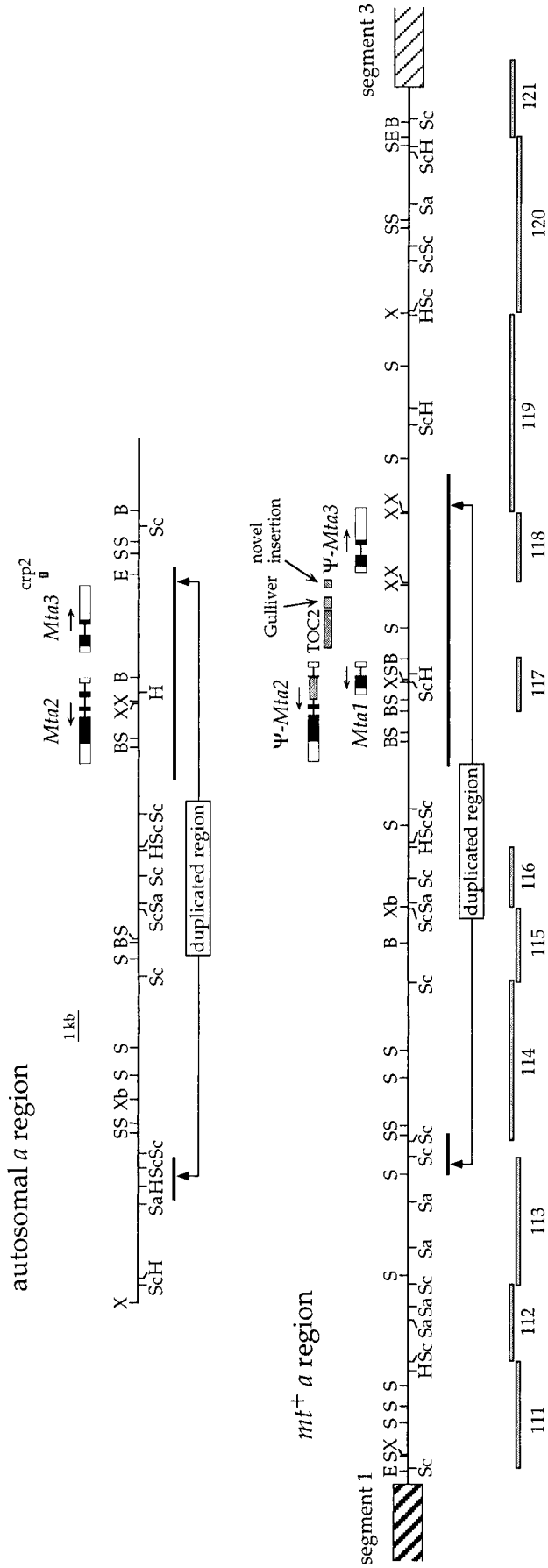

FIGURE 5.—Maps of the autosomal (top) and mt^+ (bottom) genomic regions carrying the duplicated *a* regions (bracketed by arrows; see Figure 3 for a key to restriction enzyme abbreviations). The two solid lines just under each map indicate the DNA that has been sequenced. The autosomal map shows the exon-intron structures of the *Mta2* and *Mta3* genes and the location of a crp2 repetitive element (DAY and ROCHAIX 1989). The mt^+ map shows the region between segments 1 and 3 (*cf.* Figure 1), beneath which are indicated the regions corresponding to probes 111–121. The structures of the *Mta1* gene and the Ψ-*Mta2* and Ψ-*Mta3* pseudogenes are depicted, as well as the locations of three insertions in the intergenic region between the Ψ-*Mta2* and Ψ-*Mta3* promoters.

```
  1  MRTSLWGHARPHAPPHTHPVVGYSQFPRFFPSRQFSMGCIINQLDAD

 48  LKATKLVHEQE
     VRELKLVYEEE
     ARELRRVHEEA
     VGELNRVHAEV
     VGELKRVHAEE

103  VCMLKEEVAELQRELESGTRGLRR
```

Figure 6.—Predicted sequence of the Mta1 protein, with the five 11-amino-acid long repeats aligned. The H and L residues conserved in the homologous protein from Plasmodium (Werner *et al.* 1998) are shown in boldface type.

Table 2 shows the level of homology between *Mta2* and Ψ-*Mta2* and between *Mta3* and Ψ-*Mta3*. The density of codon and noncodon differences is comparable in the two gene pairs, consistent with the possibility that the two pseudogenes were created at a similar time during *C. reinhardtii* evolution.

Insertions in the *a* region: As detailed in the discussion, the configuration of the *a* sequences in the mt^+ locus is most readily explained by proposing that the *Mta1* gene transposed into the region, thereby inactivating the resident *Mta2* gene and creating Ψ-*Mta2*. The presence of three insertions between the Ψ-*Mta2* and Ψ-*Mta3* sequences (Figure 5), which may have participated in *Mta3* inactivation, offers additional evidence of transpositional activity in the region.

The first insertion is a 1278-bp sequence related to the TOC2 element described by Day (1995). The insertion has a perfect 14-bp inverted repeat at the two ends that is identical to one of the 14-bp TOC2 inverted repeats. The 60 bp at the left end of the insertion is an 83% match to one end of TOC2, and the 26 bp at the right end is a 92% match to the other end of TOC2 and, like TOC2, the insertion has created a 7-bp target-site duplication. However, the bulk of the insertion otherwise bears little resemblance to TOC2 or to any other sequence in the database.

The second insertion is a 249-bp sequence that resembles the 12-kb Gulliver transposon (Ferris 1989): It has perfect 15-bp inverted repeats at the two ends that are a 14/15 match for the Gulliver right-end inverted repeat and creates an 8-bp target-site duplication like Gulliver. However, the sequence between the inverted repeats bears no resemblance to the limited sequences available for full-length Gulliver elements (Ferris 1989).

The third insertion is a 361-bp sequence with a direct repeat of 34 bp at each end (1-bp mismatch). There is no unambiguous target-site duplication and no homology to previously characterized Chlamydomonas transposons.

The *Ezy2* gene cluster: An obvious structural difference between the mt^+ and mt^- locus is a 16-kb DNA sequence tandemly repeated six to eight times in segment 3 of the mt^+ R domain (Figure 1). This sequence is found in the mt^- locus as a single copy, split in two, a portion resident at the end of segment 3 and the remainder resident in the C domain (Ferris and Goodenough 1994).

To determine whether gene(s) are located within the 16-kb element, an mt^+ genomic clone of the repeat unit was used to probe Northern blots. No signals were detected using vegetative or gametic samples, whereas a single 3.9-kb mRNA was detected in the 1-hr zygote sample (Figure 2F). A cDNA library generated from 1-hr zygotes was also screened with the probe, and one full-length cDNA was recovered and sequenced. An open reading frame of 3078 bp defines the unit gene, hereafter called *Ezy2* (*E*arly *z*ygote 2; Figure 7). A genomic copy was also sequenced, which showed polymorphisms in its 3′ UTR sequence to the full-length cDNA. Additional partial cDNAs were also characterized, some displaying polymorphisms to the full-length clone, suggesting that several, and perhaps all, of the *Ezy2* repeats are transcribed.

The predicted Ezy2 polypeptide is shown in Figure 8. It displays a putative 42-amino-acid chloroplast transit peptide (Figure 8, boxed): An alanine follows the initiator methionine and the N-terminal region displays a high content of valine, alanine, and serine, albeit there are fewer arginines than expected for a transit peptide (Von Heijne *et al.* 1989). The VXA predicted cleavage site (Franzen *et al.*1990) follows position 42, generating a mature polypeptide of 983 amino acids. The predicted size of Ezy2, minus the transit peptide, is 104 kD, and the predicted p*I* is 9.9, meaning that it might interact with DNA or with an acidic protein such as Ezy1 (Armbrust *et al.* 1993). However, no obvious DNA-binding or protein-binding motifs are present within the sequence, and no informative matches have been identified in the database. An intriguing feature of the sequence is that it displays a perfect internal direct repeat of 214 amino acids (Figure 8, boldface type followed by italics).

Figure 9A shows the pattern of *Ezy2* expression during zygote development as monitored by RNase protection assays. The message appears almost immediately after zygote formation, peaks at 30 min, is greatly reduced by 2 hr, and is undetectable by 4 hr into zygote development. By comparison, *Ezy1* expression peaks later (Figure 9A), as does the expression of most other zygote-specific genes (Ferris and Goodenough 1987; Uchida *et al.* 1993; Kuriyama *et al.* 1999; Suzuki *et al.* 2000).

To determine whether the bisected copy of *Ezy2* in the mt^- locus is expressed, a mating was performed between a normal *minus* strain and a *mid-1* mt^- strain transformed with the *Fus1* gene [*mid-1* mt^- *(Fus1)*]. The *mid-1* mutant, lacking a functional *Mid* gene, differentiates as *plus* and, when transformed with *Fus1*, is able to mate with *minus* gametes and form apparently normal zygotes (Ferris *et al.* 1996) that carry two copies of the bisected *Ezy2* gene in their mt^- chromosomes but no copies of the full-length *Ezy2* sequences because they lack mt^+ chromosomes. When these zygotes were subjected to RNase protection assays, no *Ezy2* expression was detected (Figure 9B), indicating that the bisected sequence is a nonfunctional gene that we henceforth designate Ψ-*Ezy2*. As a control, RNase protection was also performed using

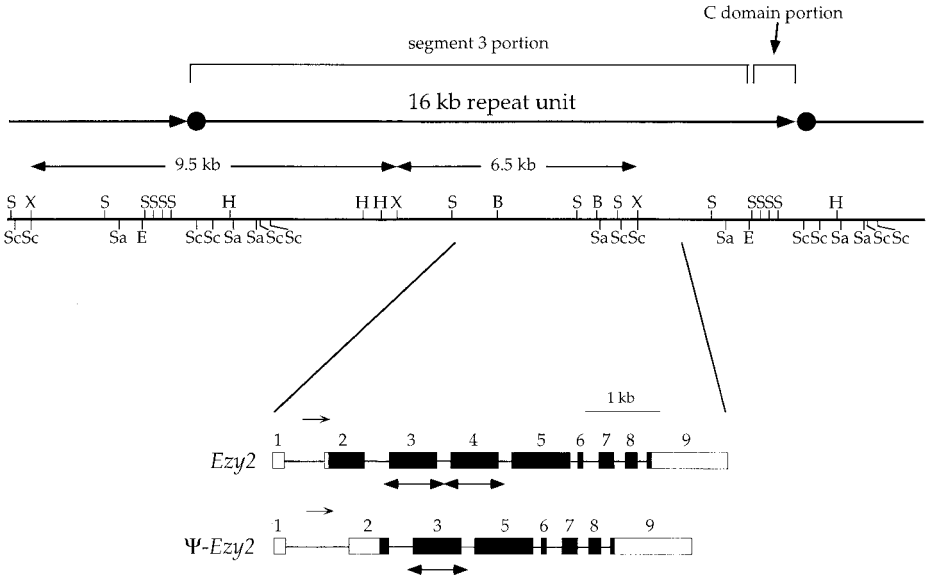

FIGURE 7.—Structure of the *Ezy2* region and comparison of *Ezy2* and Ψ-*Ezy2*. (Top) The organization of ~1.5 repeats of the 16-kb repeat unit (see Figure 3 for a key to restriction enzyme abbreviations), where a single repeat unit is indicated by the arrow with a solid circle on one end and an arrowhead on the other end. The locations of the ~6.5- and ~9.5-kb *Xho*I fragments used as probes are indicated by the double-headed arrows. Indicated at the top are the portion of the 16-kb repeat that resides in segment 3 of mt^- and the portion that resides in the C domain of mt^- (*cf.* Figures 1 and 10). (Bottom) Comparison of the *Ezy2* and Ψ-*Ezy2* gene structures. The direction of transcription is indicated; a pair of double-headed arrows shows the location and extent of the exon-intron duplication. The structure shown for the Ψ-*Ezy2* gene is hypothetical in the sense that it is no longer transcriptionally active.

the *Ezy1* sequence, a gene tandemly repeated in both the mt^+ and mt^- loci (Figure 1 and ARMBRUST *et al.* 1993), and expression was detected (Figure 9B), demonstrating that transcription of *mt*-linked genes is not generally impaired in these unusual zygotes.

The mt^+ *Ezy2* gene is ~6 kb, with a contiguous "spacer" of ~10 kb, meaning that the repeats in mt^+ segment 3 span ~100–140 kb. The gene has one intron in the 5′ UTR and seven introns in the coding region (Figure 7). The 214-amino-acid internal repeat is encoded by exons 3 and 4 (Figure 7). The first internal repeat is 829 bp and the second is 842 bp, the length differences created by three *in*sertions/*del*etions (indels) in the intervening intron. The introns are otherwise identical, and one synonymous codon difference is found between the duplicated exons. Restriction analysis indicates that the internal repeat is present in all the mt^+ *Ezy2* copies.

The genomic sequence of Ψ-*Ezy2* was also determined. Whereas the restriction maps of the mt^+ versions of *Ezy2* are very similar, the restriction maps of *Ezy2* and Ψ-*Ezy2* share few common sites (FERRIS and GOODENOUGH 1994). However, the overall sequence homology between them is sufficiently high, and the intron/exon structure sufficiently well preserved, to allow an unambiguous alignment (Figure 7). In the Ψ-*Ezy2* sequence, the spacer domain has been truncated at a downstream position, the missing portion now being located in the C domain (Figure 7). The most obvious difference between the coding regions of *Ezy2* and Ψ-*Ezy2* is that Ψ-*Ezy2* lacks exon 4 and hence the internal direct repeat (Figure 7). In addition, a frameshift at the 5′ end of Ψ-*Ezy2* shifts the location of the first candidate initiator methionine to a more downstream position (Figure 7), and numerous nucleotide differences and indels have accumulated throughout the two versions of the gene (Table 2).

Zygote development in the absence of a mt^+ locus: As noted earlier, because the *mid-1* mt^- mutant lacks a functional *Mid* gene, it expresses *plus* gametic traits; moreover, if it has been transformed with the *Fus1* gene from the mt^+ locus, it is able to fuse with mt^- gametes, generating zygotes that have two copies of the mt^- version of chromosome VI. These zygotes are apparently able to mature and germinate normally, indicating that the program for zygote development does not require genes such as *Mta1* or *Ezy2* that are exclusively encoded in the mt^+ locus.

MAVACAVAVR PLVQVAVASA VSTAAPASSK PAVKLAASAV SAVALTTVSV
SAGLLATTAV EDPRFHAADC QSRSADASAS CEDLQPSTST CTSAVRDANR
PTRRVRRSGS KAQRRGSTTL TASVPSMAAA VVLPPKIALR RRHRLRLRAG
HSATAAATDK TPREQPD**KPA ALPEDLLPAD ATSTSSTGKI SSAAVCCGLL**
AHCSAAQLHA ILCGLVQAVA SSSVKGNNRK LLLGSKLRKL LEGVGVAPAN
GKAYTAADVA ALSGPKLERL RATLKSQPGL LLWFLLFTAP AKLQALQAAL
LPGGAGDRSF EEWRAAIDAV AGSGHEQLAA AQEVRGRQSA CVEGSTAGNT
ATTATITTTN NNPASHGGVY TALTGTEVTG K*KPAALPEDL LPADATSTSS*
TGKISSAAVC CGLLAHCSAA QLHAILCGLV QAVASSSVKG NNRKLLLGSK
LRKLLEGVGV APANGKAYTA ADVAALSGPK LERLRATLKS QPGLLLWFLL
FTAPAKLQAL QAALLPGGAG DRSFEEWRAA IDAVAGSGHE QLAAAQEVRG
*RQSACVEGST AGNTATTATI TTTNNNPASH GGVYTALTGT EVTGK*AAANK
DLSRTRTTSH RNRCVSESGS TRNKSRSSSS RSSSTHSVEY AEPKAGCSQP
AATVPGCVPE IISAAIPPLA PLALHIRRAI VKELLEARPP GWNTFLYSWL
QAAGLSEFLP ANGTCRMYMA DRKQLVLRVG AMREEQVDAF LTCMCKAHGH
STWLARYLHM LGPEVSQLLS QGRYSDELLA ALRAAGQKTL ADAVMEHFWG
RDPDPEDSEA GEMDVKPWAE RLGLLRFDML AEQLRLPPNA DGSVKNFSNG
LVFKVDPLEV WSKYTDGEPS AGALSGMRAT DKEARDKQVK QLRGVPLLYL
WRIGGRVVYV GMSGGWVKGR RIARYLAEGP GFSESSKMLP WLTAIDEGKE
IELRVITLEG LKALEGMSEG MSEEEVQKKV QKKVKELEKH FLCHVDCPCN
KVNNGSYRVE TPRQASWTNS RRSTR

FIGURE 8.—Predicted sequence of the Ezy2 protein. The putative chloroplast transit peptide is boxed. The first internal 214-amino-acid repeat is in boldface type, and the second is in italics.

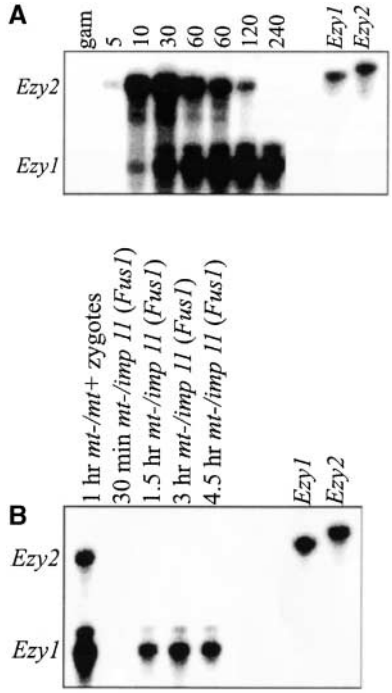

FIGURE 9.—RNase protection analysis of *Ezy2* gene expression. (A) Total RNA was extracted from gametes and from zygotes 5, 10, 30, 60, 120, or 240 min after zygote formation and was hybridized with *Ezy1* or *Ezy2* antisense RNA probes. The *Ezy1* probe is 176 nucleotides (nt), and the protected fragment is 118 nt. The *Ezy2* probe is 197 nt and the protected fragment is 172 nt. (B) RNase protection analysis of *Ezy1* and *Ezy2* message levels in either wild-type zygotes or zygotes resulting from a cross between mt^- and *mid-1 (Fus1)* gametes. Total RNA was isolated from wild-type mt^+/mt^- zygotes 1 hr after the gametes were mixed or from the mt^- *mid-1 (Fus1)* zygotes at the indicated intervals. The lower expression of the *Ezy1* message in the mutant zygotes is presumably due to the fact that cell fusion is not efficient in these matings.

We went on to ask whether the uniparental transmission of *plus* cpDNA is affected in these zygotes. Table 3 compares the transmission patterns of chloroplast markers in control crosses and crosses in which a *mid-1* mt^- *(Fus1)* strain served as the *plus* parent. Inheritance of chloroplast traits is seen to be biparental in the crosses involving the strain lacking an mt^+ chromosome.

DISCUSSION

Coding capacity of the *MT* locus: The rearranged R domain of the *C. reinhardtii MT* locus as well as the flanking T and C domain sequences that are under recombinational suppression are shown to contain genes that are expressed throughout the life cycle of the organism as well as genes expressed exclusively during the gametic or the zygotic phases of the life cycle. Although a comparable transcription map has not yet been generated for other regions of the *C. reinhardtii* genome, this distribution of genes is what one would expect if an ordinary chromosome had undergone large-scale rearrangements and had also gained a few gene sequences in one homolog but not the other. Large-scale rearrangements are found in the mouse *T* locus, which includes genes affecting male fertility (SILVER 1985; LYON *et al.* 2000), and in the self-incompatibility loci of Brassica plants (CASSELMAN *et al.* 2000; KUSABA *et al.* 2001), and genes-without-homologs characterize mating-type loci in the fungi (KRONSTAD and STABEN 1997; BADRANE and MAY 1999) and XY chromosome pairs in mammals (LAHN and PAGE 1999). Unusual architecture therefore appears to be a common feature of sex-related chromosomal domains.

A particular goal of this study was to ascertain whether the four large segments of rearranged DNA in the R domain contain active genes or are instead noncoding structural elements, as is the case, for example, for most of the mammalian Y chromosome (LAHN and PAGE 1997). Numerous active genes were in fact identified throughout the R domain (Figure 1 and Table 1). One of the genes (*Pdk*) encodes pyruvate dehydrogenase kinase, an enzyme that plays a key role in controlling pyruvate dehydrogenase activity and hence the TCA cycle and cellular respiration (ZOU *et al.* 1999). A second (*GdcH*) encodes glycine decarboxylase complex subunit H, an enzyme of the photorespiration pathway (OLIVER 1994; SRINIVASAN and OLIVER 1995). Two other sequenced genes (*Pr46a* and *Pr46b*), while of unknown function, have well-conserved homologs in several multicellular eukaryotes. Therefore, the R domain of chromosome VI appears to have maintained many, and perhaps all, of its prior genetic activities while having been subjected to numerous local rearrangements and insertion/deletion events.

It is widely assumed that one of the functions of meiotic recombination is to promote genomic integrity, and it has been demonstrated that chromosomes prevented from engaging in meiotic recombination are subject to deterioration, a model for the ontogeny of XY differentiation (CHARLESWORTH 1991; RICE 1994). One would therefore not expect important enzymes such as quinolinate synthetase, glycine decarboxylase, hydroxyethylthiazole kinase, and pyruvate dehydrogenase kinase, and important transcriptional regulators such as Rb, to be encoded in genomic regions that are under heavy recombinational suppression. Presumably any costs incurred by this suppression are offset by the advantage it confers, but the nature of the advantage has yet to be determined.

***MT*-unique sequences:** We also examined closely six regions of the *MT* locus that are found in one chromosome but not the other; these are hereafter referred to as *MT*-unique sequences. We were unable to detect any genes in two of these—region *b* in mt^+ and region *e* in mt^- (see data at http://www.genetics.org/supplemental)—albeit it is of interest that region *b* is duplicated, in inverted orientation, at a site 1 cM telomere-proximal to the mt^+ locus (FERRIS and GOODENOUGH 1994), yet another example of autosome/*MT* duplication. The re-

TABLE 3

Uniparental inheritance of chloroplast markers

		Individual progeny			Zygotes[a]			
plus parent	*minus* parent	UP+	UP−	UP− (%)	UP+	BP	UP−	Exceptional (%)
		Control cross ($mt^+ \times mt^-$)						
CC-118 sr^r	CC-124	87 sr^r	0 sr^s	0	22	0	0	0
		mid-1 mt⁻ (Fus1) × *mt⁻* crosses						
mid-1 (Fus1)	CC-421	53 spr^s	42 spr^r	44	9	5	7	57
B32	CC-1952	40 spr^r	74[c] spr^s	65	4	15	13	88

Individual progeny were scored as UP+ if they have the drug-resistance phenotype of the *plus* parent or UP− if they have that of the *minus* parent (of if they were among the few progeny that were clearly a mixture of sensitive and resistant cells). Zygotes were scored as UP+ if all the individual progeny from that zygote were UP+, as UP− if all were UP−, and biparental (BP) if there were progeny of both types. Zygotes that are BP or UP− are considered exceptional.

[a] Only zygotes with three or four surviving progeny are included.

[b] Includes five progeny containing both spr^r and spr^s cells.

maining four *MT*-unique sequences appear to contain one active gene apiece. Each is restricted in expression to the gametic phase of the life cycle; two are *plus* specific and two are *minus* specific.

1. Region *a* in mt^+ contains the gene *Mta1* that is expressed in *plus* gametes only. The Mta1 protein is predicted to contain a leucine-histidine zipper and is of unknown function.
2. Region *c* in mt^+ contains the *Fus1* gene, encoding the Fus1 protein, that is expressed in *plus* gametes only and is necessary for *plus*-mediated gametic cell fusion (Ferris *et al.* 1996).
3. Region *d* in mt^- contains the *Mtd1* gene that is expressed in *minus* gametes only. The predicted Mtd1 gene product is a putative triple-span membrane protein with putative extracellular N-glycosylation sites. When mt^+ gametes are transformed with the *Mid* gene, which causes them to differentiate as *minus*, their flagellar agglutination is strong but their cell fusion is very slow and erratic. Since such gametes lack the *Mtd1* gene, there is a pleasing symmetry to the possibility that regions *c* and *d* might contain genes *Fus1* and *Mtd1* that code for *plus* and *minus* cell-fusion proteins, respectively. The Mtd1 sequence shows no homology to known membrane-fusion motifs, so if it proves to participate in membrane fusion it may do so by a novel mechanism.
4. Region *f* in mt^- contains the *Mid* gene that is expressed in *minus* gametes only. The Mid protein is necessary for *minus* gametic differentiation (Ferris and Goodenough 1997). Whereas there would be a pleasing symmetry in the postulate that the Mta1 protein is necessary for *plus* gametic differentiation, this is ruled out by the ability of *mid-1* and *mid-2* mutants to differentiate as *plus* gametes in the absence of an *Mta1* gene.

In addition to these four genes, the *Ezy2* gene is *MT*-unique as well, being expressed from the mt^+ locus only. It differs from the four genes above in three respects: It is present in multiple tandem copies; its expression is initiated in the zygote rather than in the gamete; and it is not strictly unique to the mt^+ locus in that a nonexpressed *Ezy2* pseudogene is located in the mt^- locus.

Codon bias: The first two genes to be sequenced from the *C. reinhardtii MT* locus were *Fus1* (Ferris *et al.* 1996) and *Mid* (Ferris and Goodenough 1997), and both had the surprising property of lacking the codon bias found in all other *C. reinhardtii* genes, generating the suggestion that bias might be relaxed because these genes both reside in the R domain and/or because both lack homologs. Table 1 documents that neither suggestion is generally applicable: The R-domain genes identified in this study all show moderate to strong codon bias (B value) and a high percentage of GC, including *Mta1* and *Mtd1*, which have no homologs, and *Ezy2*, which no longer has a functional homolog. Therefore, the absence of bias in *Fus1* and *Mid* remains unexplained, although it may indicate that they have been without homologs longer than the other genes (Kliman and Hey 1993).

Chloroplast DNA inheritance: During the first 2 hr of zygote maturation in *C. reinhardtii*, cpDNA derived from the mt^- parent is normally degraded by nuclease digestion whereas cpDNA from the mt^+ parent is preserved and later selectively replicated (Umen and Goodenough 2001a), resulting in the uniparental-*plus* pattern of inheritance of chloroplast-encoded traits (Armbrust 1998). It has been postulated that this system is analogous to modification/restriction systems in bacteria, with the *plus* cpDNA being selectively "protected" by methylation so that it resists cutting by methylation-sensitive restriction enzymes in the zygote (Sager and Kitchin 1975). However, recent studies do not support such a model (Umen and Goodenough 2001a), and the molecular basis for uniparental-*plus* inheritance awaits elucidation.

Matagne and Mathieu (1983) observed that when

heterozygous diploid (mt^+/mt^-) *minus* strains were crossed with either haploid *plus* or homozygous diploid (mt^+/mt^+) *plus* strains, cpDNA transmission was biparental. These findings were interpreted to indicate that "protection" of *plus* cpDNA in the mt^+/mt^- parent is dependent on the presence of the mt^+ locus and is not subject to "*minus* dominance" (*i.e.*, is not Mid-repressible). We show here that the *Mta1* gene is restricted in expression to mt^+ gametes and is not Mid-repressible (Figure 2E). However, our results would seemingly argue against a role for *Mta1*, or any other gene in the mt^+ locus, in cpDNA protection since, in the absence of a mt^+ chromosome (*e.g.*, in *mid-1* mt^- *(Fus1)* × mt^- crosses), zygotes give rise to viable meiotic progeny. If both *plus* and *minus* cpDNA were unprotected and hence destroyed in the early zygote, the cross would presumably be lethal, as is indeed the case in a related system (Vanwinkle-Swift *et al.* 1994).

The *mid-1* mt^- *(Fus1)* × mt^- cross is not lethal, but neither is it normal: cpDNA is inherited biparentally (Table 3), suggesting that the missing mt^+ chromosome is somehow necessary for the selective destruction of *minus* cpDNA in the zygote. For example, if the mt^+-encoded Ezy2 protein participates in cpDNA destruction and is selectively targeted to *minus* chloroplasts in the zygote (perhaps because *minus* chloroplasts carry specific receptors for Ezy2 translocation; *cf.* Bauer *et al.* 2000), then biparental inheritance would be expected to occur in the absence of Ezy2.

Taken together, the results available at present are best explained by proposing that the mt^+ locus encodes *both* a protection function *and* a destruction function, with Mta1 being a candidate participant in protection and Ezy2 in destruction. Both of these functions would be operative in the heterozygous-diploid crosses of Matagne and Mathieu (1983), generating two sets of protected genomes and hence biparental inheritance. By contrast, neither set of functions would be operative in our crosses, which would also result in biparental inheritance because neither set of unprotected genomes would be destroyed.

Evolutionary history of the *MT* locus: A common way to model the evolution of separate sexes (heterothallism, dioecy) is to start with a self-fertile (homothallic, monoecious) ancestor and propose steps that would lead to self-sterility (*e.g.*, Charlesworth 1991). A homothallic lineage ancestral to *C. reinhardtii* can most simply be thought of as having a *Mid* gene in chromosome VI that switched "on" in some cells and "off" in others, the former cells expressing *minus*-specific genes and hence differentiating as *minus* gametes, and the latter cells expressing *plus*-specific genes and differentiating as *plus* gametes. Indeed, this is the inferred pattern of gene expression in the distantly related homothallic species *C. monoica* (Vanwinkle-Swift *et al.* 1998). The loss of *Mid* from a copy of chromosome VI would then generate a self-sterile *plus*-only clone carrying a proto-mt^+ locus, while the loss of the off switch from the *Mid* gene in another copy of chromosome VI would generate a self-sterile *minus*-only clone carrying a proto-mt^- locus.

To model the subsequent "invasion" of the homothallic population by these two chromosomes, one can invoke the benefits of outcrossing as driving the process. Alternatively, or in addition, one can invoke positive selection for advantageous genes linked to the proto-*mt* loci and propose that the linkage would come to be buttressed by recombinational suppression (Charlesworth 1991; Rice 1994; Trickett and Butlin 1994; but see Filatov *et al.* 2000). One suggestion along these lines has been that linkage disequilibrium came to preserve an adaptive association between mating type and genes involved in organelle DNA inheritance (Hurst 1992; Hurst and Hamilton 1992).

We can now consider possible origins of the *MT*-unique genes, using *Fus1* as an example. There are two possibilities: Either *Fus1* originally happened to reside in the proto-*MT* region of chromosome VI and was subsequently lost from the proto-mt^- chromosome or it was originally autosomal and then moved into the proto-mt^+ locus, subsequently losing its autosomal representation and its *Mid*-repressibility. In either case, once any mating-related gene like *Fus1* became *MT*-unique, it would become dependent on its *MT*-linkage for correct expression in *plus* or *minus* gametes. Thus, the acquisition of one or more *MT*-unique gametogenesis genes would lead to a selective advantage for chromosomal rearrangements or other changes that (further) suppress recombination in the region, thereby assuring that a gene like *Fus1* is expressed in mt^+ gametes and not expressed in mt^- gametes.

A mating-related gene could originate in or move into an *MT* locus by chance, and the loss of additional gene representation and of *Mid* regulation could also occur by chance. Alternatively, there may be some selective advantage to a *cis*-configuration of gametogenesis genes, as opposed to regulating their expression *in trans*. Since our results indicate that such "gene acquisition" events have occurred several times during the evolutionary history of the *MT* locus in *C. reinhardtii*, a selective advantage is suggested, but its nature remains to be identified.

Mutational profile of the *MT* locus: The most striking feature of the *MT* locus is its unusual chromosomal organization (Ferris and Goodenough 1994). The present study provides additional examples of unconventional configurations.

As summarized in Figure 5, the *a* region, present in the mt^+ locus and absent from the mt^- locus, is a sequence that is duplicated in an autosome and flanked by DNA that carries no identified genes. The autosomal copies of the *a*-region genes (the gamete-specific *Mta2* and the housekeeping *Mta3*) are functional, whereas their mt^+ counterparts are pseudogenes (Ψ-*Mta2* and Ψ-*Mta3*). Of particular interest is the gamete-specific *Mta1* gene in the mt^+ locus, which co-opts the upstream regulatory elements and the first nine codons of an *Mta2* sequence

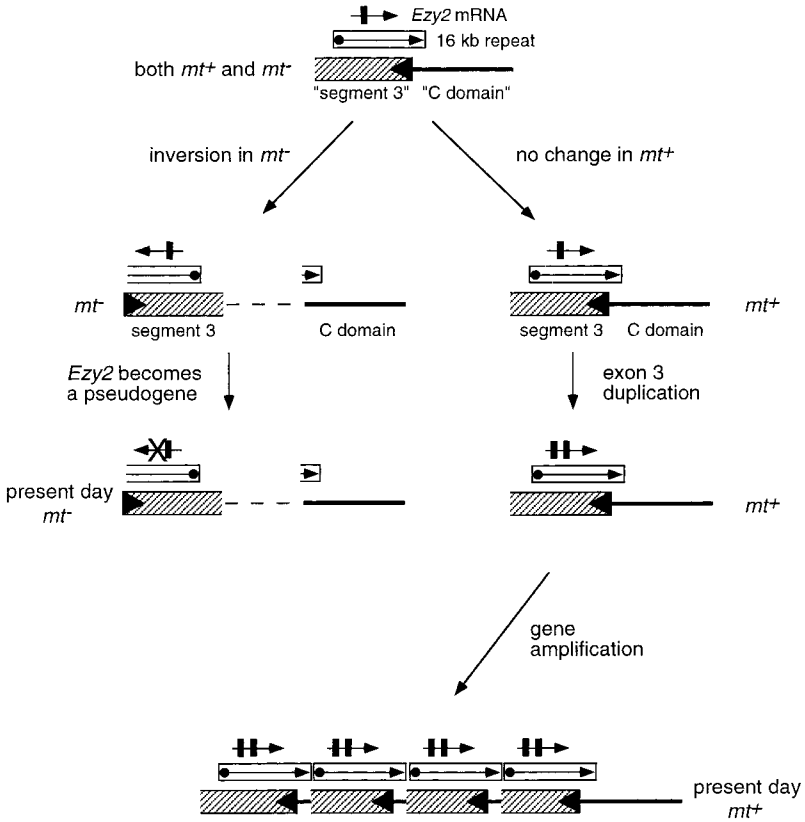

FIGURE 10.—Postulated sequence for the evolution of *Ezy2* genes. The original *Ezy2*, present in both mating types, is depicted as a single-copy gene flanked by an untranscribed "spacer" and containing one copy of the now duplicated exon 3. A double-stranded break occurred within the spacer region of the mt^- copy, allocating a portion of the spacer to what is now the mt^- C domain (boxed arrowhead) and the rest of the gene/spacer to what is now the mt^- segment 3. [Segment 3 is presently inverted and is also separated from the C domain by segment 4 (Figure 1); the timing of these rearrangements vis-à-vis the evolution of the *Ezy2* sequences is unknown.] The loss of the C-domain spacer sequences may have participated in rendering the mt^- copy of the gene transcriptionally inactive and hence a pseudogene, or transcriptional inactivation may have occurred for other reasons: For a zygote-specific gene linked to mating type, loss of gene activity in one of the two loci may be difficult to select against. The mt^+ copy, meanwhile, underwent a duplication of exon 3 and then subsequently underwent an expansion in copy number, perhaps in part to make up for the loss of the mt^- copy. An alternative scenario would propose that *Ezy2* was originally a multigene family in both mating types, each gene containing an unduplicated exon. In mt^-, the subsequent chromosomal rearrangements deleted all but one copy, which subsequently became a pseudogene. In mt^+, the exon duplication occurred in one copy and then spread by concerted evolution (SWANSON and VAQUIER 1998).

and then diverges into a unique open reading frame (ORF), the resulting gene being a chimera (Figure 5).

The most likely scenario for the generation of this chimera is to propose that the *Mta1* sequence inserted into a preexisting *Mta2* gene, thereby capturing a gamete-specific promoter, which is analogous to the acquisition of a testis-specific promoter by the *Cdic* gene in *Drosophila melanogaster* (NURMINSKY *et al.* 1998). This scenario leaves open many questions: Where did the *Mta1* sequence come from? Was it autosomal or *MT*-linked and what were its original upstream sequences? Did an intact *Mta2* gene originally exist in the mt^+ locus, which was then disrupted by the *Mta1* transposition event? Or was the chimeric sequence constructed in an autosome, perhaps in a duplicated copy of *Mta2*, and then transposed to the mt^+ locus?

The *GdcH*, *Pr46a*, and *Pr46b* genes in segment 3 illustrate a different kind of unusual gene overlap. As summarized in Figure 3, each of these genes overlaps one of the others at its 3′ end, but none of these overlaps have led to gene inactivation since all three are transcribed. Nothing is known about how these relationships were established, but, given the high density of rearrangements in the *MT* locus, it is possible that the three genes were once separated and were subsequently brought together. This is, to our knowledge, the first report of nuclear gene overlap in *C. reinhardtii*.

The major rearrangements involving segments 1–4 have generated two mutations characterized in this study. First, the distal portion of the 3′ UTR of the *Pdk* gene, located at one edge of segment 2, contains region-*b* sequences in the mt^+ locus and completely different segment 1 sequences in the mt^- locus (Figure 1). Presumably one of these sequences represents the original 3′ UTR and the other was created by rearrangement; it is not known whether these differences affect the properties of the two gene transcripts. Second, rearrangements involving segment 3 of the mt^- locus have disrupted the *Ezy2* gene. The large number of differences between *Ezy2* and Ψ-*Ezy2* compared to other gene/pseudogene pairs in the *MT* locus (Table 2) suggests that this event occurred in the more distant past.

The *Ezy2* configurations are particularly intriguing in that they entail four different kinds of alterations: (1) rearrangement of gene order; (2) inactivation of the gene in the mt^- locus; (3) endoduplication of an exon in the mt^+ locus gene; and (4) tandem duplication of the endoduplicated gene to generate six to eight copies. Figure 10 presents a possible scenario for the sequence of these three events, with details given in the legend.

Several highly expressed autosomal zygote-specific genes have previously been found to exist as near-neighbor duplicates, including two cases in which both copies are functional (UCHIDA *et al.* 1999; SUZUKI *et al.* 2000) and one case in which one copy is now a pseudogene (MATTERS and GOODENOUGH 1992). Two additional ex-

amples of apparent near-neighbor duplicates of zygote-specific genes—detected by Pr 72/74 and by Pr 100 (Table 1)—have been found in this study. However, the long tandem iteration of *Ezy2* genes and the nearby tandem cluster of zygote-specific *Ezy1* genes in both chromosomes (Figure 1) clearly represent a distinctive phenomenon and one that appears to be a recurring theme in sexual evolution. In *D. melanogaster*, for example, a recent 10-fold tandem iteration of a sperm-specific gene is found in the X chromosome (Nurminsky *et al.* 1998), and tandemly repeated genes are the rule in the human Y chromosome (reviewed in Lahn and Page 1997). Indeed, features of the *AZF* region of the human Y (Saxena *et al.* 1996) offer striking parallels to the *MT* locus. *AZF* contains multiple copies, >99% identical in sequence, of a gene called *DAZ* (*D*eleted in *A*zoospermia), an RNA-binding protein essential for male fertility. The *DAZ* sequence is found as well in human chromosome 3, where expression is restricted to the germ cells of both sexes. During primate evolution, a copy of this autosomal gene transposed to the Y, where one of its exons underwent internal amplification, after which the modified gene itself underwent amplification.

The data reported here, combined with previous studies, reveal the *MT* locus to be an unusual and dynamic region of the *C. reinhardtii* genome, harboring translocations, inversions, large indels, genes without homologs, genes that transpose (Ferris and Goodenough 1997), tandem gene duplications, gene inactivation events, and, in two genes, unusual codon bias. Moreover, the sex-related genes in the locus have been shown to be undergoing rapid evolution between species (Ferris *et al.* 1997). And yet, despite these anomalies, the locus continues to encode large numbers of housekeeping genes that presumably occupied this region of chromosome VI long before it took on its modern configuration and novel functions. Since most of the *C. reinhardtii* life cycle is carried out in the haploid state, the presence of these presumably essential genes may keep a selective brake on what would, in a predominantly diploid organism, be a far more extensive, Y chromosome-like reconfiguration of the region. If so, then the *MT* locus may offer a unique opportunity to observe sex-chromosome evolution in progress.

We thank Christoph Beck for providing the *mid-2* deletion strain, the Kazusa Institute for sending several cDNA clones, and Kosuke Shimogawara for providing the sequence of the *Thi10* gene. We also thank Chunsheng Luo, Eileen Westphale, and Linda Small for excellent technical assistance. This study was supported by grants from the National Science Foundation (MCB-9904667) and the U.S. Public Health Services (GM-26150).

LITERATURE CITED

Armbrust, E. V., 1998 Uniparental inheritance of chloroplast genomes, pp. 93–113 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*, edited by J.-D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Armbrust, E. V., P. J. Ferris and U. W. Goodenough, 1993 A mating type linked gene cluster expressed in Chlamydomonas participates in the uniparental inheritance of the chloroplast genome. Cell **74:** 801–811.

Armbrust, E. V., A. Ibrahim and U. W. Goodenough, 1995 A mating type linked mutation that disrupts the uniparental inheritance of chloroplast DNA also disrupts cell-size control in *Chlamydomonas*. Mol. Biol. Cell **6:** 1807–1818.

Asamizu, E., Y. Nakamura, S. Sato, H. Fukuzawa and S. Tabata, 1999 A large scale structural analysis of cDNAs in a unicellular green alga, *Chlamydomonas reinhardtii*. I. Generation of 3433 non-redundant expressed sequence tags. DNA Res. **6:** 363–373.

Asamizu, E., K. Miura, K. Kucho, Y. Inoue, H. Fukuzawa *et al.*, 2000 Generation of expressed sequence tags from low-CO_2 and high-CO_2 adapted cells of *Chlamydomonas reinhardtii*. DNA Res. **7:** 305–307.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman *et al.*, 1989 *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.

Badrane, H., and G. May, 1999 The divergence-homogenization duality in the evolution of the *b1* mating type gene of *Coprinus cinereus*. Mol. Biol. Evol. **16:** 975–986.

Bauer, J., K. Chen, A. Hiltbunner, E. Wehrli, M. Eugster *et al.*, 2000 The major protein import receptor of plastids is essential for chloroplast biogenesis. Nature **403:** 203–207.

Casselman, A. L., J. Vrebalov, J. A. Conner, A. Singhal, J. Giovannoni *et al.*, 2000 Determining the physical limits of the *Brassica S* locus by recombinational analysis. Plant Cell **12:** 23–34.

Charlesworth, B., 1991 The evolution of sex chromosomes. Science **251:** 1030–1033.

Day, A., 1995 A transposon-like sequence with short terminal inverted repeats in the nuclear genome of *Chlamydomonas reinhardtii*. Plant Mol. Biol. **28:** 437–442.

Day, A., and J.-D. Rochaix, 1989 Characterization of transcribed dispersed repetitive DNAs in the nuclear genome of *Chlamydomonas reinhardtii*. Curr. Genet. 1**6:** 165–176.

Devereux, J., P. Haeberli and O. Smithies, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

Ebersold, W. T., 1967 *Chlamydomonas reinhardtii*: heterozygous diploid strains. Science **157:** 447–449.

Ferris, P. J., 1989 Characterization of a Chlamydomonas transposon, *Gulliver*, resembling those in higher plants. Genetics **122:** 363–377.

Ferris, P. J., 1995 Localization of the *nic-7*, *ac-29*, and *thi-10* genes within the mating-type locus of *Chlamydomonas reinhardtii*. Genetics **141:** 543–549.

Ferris, P. J., and U. W. Goodenough, 1987 Transcription of novel genes, including a gene linked to the mating-type locus, induced by *Chlamydomonas* fertilization. Mol. Cell. Biol. **7:** 2360–2366.

Ferris, P. J., and U. W. Goodenough, 1994 The mating-type locus of Chlamydomonas reinhardtii contains highly rearranged DNA sequences. Cell **76:** 1135–1145.

Ferris, P. J., and U. W. Goodenough, 1997 Mating type in Chlamydomonas is specified by *mid*, the *minus*-dominance gene. Genetics **146:** 859–869.

Ferris, P. J., J. P. Woessner and U. W. Goodenough, 1996 A sex recognition glycoprotein is encoded by the *plus* mating-type gene *fus1* of *Chlamydomonas reinhardtii*. Mol. Biol. Cell **7:** 1235–1248.

Ferris, P. J., C. Pavlovic, S. Fabry and U. W. Goodenough, 1997 Rapid evolution of sex-related genes in *Chlamydomonas*. Proc. Natl. Acad. Sci. USA **94:** 8634–8639.

Ferris, P. J., J. P. Woessner, S. Waffenschmidt, S. Kilz, J. Drees *et al.*, 2001 Glycosylated polyproline II rods with kinks as a structural motif in plant hydroxyproline-rich glycoproteins. Biochemistry **40:** 2978–2987.

Filatov, D. A., F. Monéger, I. Negrutiu and D. Charlesworth, 2000 Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. Nature **404:** 388–390.

Finst, R. J., P. J. Kim, E. R. Griffis and L. M. Quarmby, 2000 Fa1p is a 171 kDa protein essential for axonemal microtubule severing in *Chlamydomonas*. J. Cell Sci. **113:** 1963–1971.

Franzen, L.-G., J.-D. Rochaix and G. Von Heijne, 1990 Chloroplast transit peptides from the green alga *Chlamydomonas reinhardtii*

share features with both mitochondrial and higher plant chloroplast presequences. FEBS Lett. **260:** 165–168.

Gillham, N. W., 1969 Uniparental inheritance in *Chlamydomonas reinhardtii*. Am. Nat. **103:** 355–388.

Goodenough, U. W., P. A. Detmers and C. Hwang, 1982 Activation for cell fusion in *Chlamydomonas*: analysis of wild-type gametes and nonfusing mutants. J. Cell Biol. **92:** 378–386.

Goodenough, U. W., E. V. Armbrust, A. M. Campbell and P. J. Ferris, 1995 Molecular genetics of sexuality in *Chlamydomonas*. Annu. Rev. Plant Physiol. Plant Mol. Biol. **46:** 21–44.

Harris, E. H., 1989 *The Chlamydomonas Sourcebook*. Academic Press, San Diego.

Hurst, L. D., 1992 Intragenomic conflict as an evolutionary force. Proc. R. Soc. Lond. Ser. B **248:** 135–140.

Hurst, L. D., and W. D. Hamilton, 1992 Cytoplasmic fusion and the nature of sexes. Proc. R. Soc. Lond. Ser. B **247:** 189–194.

Hwang, C. J., B. C. Monk and U. W. Goodenough, 1981 Linkage of mutations affecting *minus* flagellar membrane agglutinability to the *mt*-mating-type locus of *Chlamydomonas*. Genetics **99:** 41–47.

Ichihara, S., T. Suzuki, M. Suzuki and S. Mizushima, 1986 Molecular cloning and sequencing of the *sppa* gene and characterization of the encoded protease IV, a signal peptide peptidase, of *Escherichia coli*. J. Biol. Chem. **261:** 9405–9411.

Kang, J.-T., and M. W. Fawley, 1997 Variable (CA/GT)n simple sequence repeat DNA in the alga *Chlamydomonas*. Plant Mol. Biol. **35:** 943–948.

Kirk, M. M., and D. J. Kirk, 1985 Translational regulation of protein synthesis, in response to light, at a critical stage of Volvox development. Cell **41:** 419–428.

Kliman, R. M., and J. Hey, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 1239–1258.

Kronstad, J. W., and C. Staben, 1997 Mating type in filamentous fungi. Annu. Rev. Genet. **31:** 245–276.

Kuriyama, H., H. Takano, L. Suzuki, H. Uchida, S. Kawano *et al.*, 1999 Characterization of *Chlamydomonas reinhardtii* zygote-specific cDNAs that encode novel proteins containing ankyrin repeats and WW domains. Plant Physiol. **119:** 873–884.

Kurvari, V., N. V. Grishin and W. J. Snell, 1998 A gamete-specific, sex-limited homeodomain protein in *Chlamydomonas*. J. Cell Biol. **143:** 1971–1980.

Kusaba, M., K. D. Dwyer, J. Hendershot, J. B. Nasrallah and M. E. Nasrallah, 2001 Self-incompatibility in the genus Arabidopsis: characterization of the *S* locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. Plant Cell **13:** 627–643.

Lahn, B. T., and D. C. Page, 1997 Functional coherence of the human Y chromosome. Science **278:** 675–680.

Lahn, B. T., and D. C. Page, 1999 Four evolutionary strata on the human X chromosome. Science **286:** 964–967.

Long, M., and J. H. Gillespie, 1991 Codon usage divergence of homologous vertebrate genes and codon usage clock. J. Mol. Evol. **32:** 6–15.

Lupas, A., 1996 Prediction and analysis of coiled-coil structures. Methods Enzymol. **266:** 513–525.

Lyon, M. F., J. C. Schimenti and E. P. Evans, 2000 Narrowing the critical regions for mouse *t* complex transmission ratio distortion factors by use of deletions. Genetics **155:** 793–801.

Magni, G., A. Amici, M. Emanuelli and N. Raffaelli, 1999 Enzymology of NAD^+ synthesis. Adv. Enzymol. **73:** 135–163.

Martin, N. C., and U. W. Goodenough, 1975 Gametic differentiation in *Chlamydomonas reinhardtii*. I. Production of gametes and their fine structure. J. Cell Biol. **67:** 587–605.

Matagne, R. F., and D. Mathieu, 1983 Transmission of chloroplast genes in triploid and tetraploid zygospores of *Chlamydomonas reinhardtii*: roles of mating-type gene dosage and gametic chloroplast DNA content. Proc. Natl. Acad. Sci. USA **80:** 4780–4783.

Matsuda, Y., T. Saito, T. Umemoto and Y. Tsubo, 1988 Transmission patterns of chloroplast genes after polyethylene glycol-induced fusion of gametes in non-mating mutants of *Chlamydomonas reinhardtii*. Curr. Genet. **14:** 53–58.

Matters, G. L., and U. W. Goodenough, 1992 A gene/pseudogene tandem duplication encodes a cysteine-rich protein expressed during zygote development in *Chlamydomonas reinhardtii*. Mol. Gen. Genet. **232:** 81–88.

Minami, S. A., and U. W. Goodenough, 1978 Novel polypeptide synthesis induced by gametic cell fusion in *Chlamydomonas reinhardi*. J. Cell Biol. **77:** 165–181.

Naver, H., P. J. Ferris and J.-D. Rochaix, 2000 The *ac29* protein is involved in the biogenesis of the light harvesting complex I and II. Abstracts, Ninth International Conference on the Cell and Molecular Biology of Chlamydomonas, Noordwijkerhout, The Netherlands.

Nurminsky, D. I., M. V. Nurminskaya, D. De Aguiar and D. L. Hartl, 1998 Selective sweep of a newly evolved sperm-specific gene in Drosophila. Nature **396:** 572–575.

Oliver, D. J., 1994 The glycine decarboxylate multienzyme complex from plant mitochondria. Annu. Rev. Plant Physiol. Plant Mol. Biol. **45:** 323–337.

Remacle, C., and R. F. Matagne, 1998 Mitochondrial genetics, pp. 661–674 in *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*, edited by J.-D. Rochaix, M. Goldschmidt-Clermont and S. Merchant. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Rice, W. R., 1994 Degeneration of a nonrecombining chromosome. Science **263:** 230–232.

Sager, R., and R. Kitchin, 1975 Selective silencing of eukaryotic DNA. Science **189:** 426–432.

Saxena, R., L. G. Brown, T. Hawkins, R. K. Alagappan, H. Skaletsky *et al.*, 1996 The *DAZ* gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified, and pruned. Nat. Genet. **14:** 292–299.

Schauser, L., A. Roussis, J. Stiller and J. Stougaard, 1999 A plant regulator controlling development of symbiotic root nodules. Nature **402:** 191–195.

Silver, L. M., 1985 Mouse *t* haplotypes. Annu. Rev. Genet. **19:** 179–208.

Smyth, R. D., G. W. Martinek and W. T. Ebersold, 1975 Linkage of six genes in *Chlamydomonas reinhardtii* and the construction of linkage test strains. J. Bacteriol. **124:** 1615–1617.

Srinivasan, R., and D. J. Oliver, 1995 Light-dependent and tissue-specific expression of the H-protein of the glycine decarboxylase complex. Plant Physiol. **109:** 161–168.

Sundberg, E., J. G. Slagter, I. Fridborg, S. P. Cleary, C. Robinson *et al.*, 1997 *ALBINO3*, an Arabidopsis nuclear gene essential for chloroplast differentiation, encodes a chloroplast protein that shows homology to proteins present in bacterial membranes and yeast mitochondria. Plant Cell **9:** 717–730.

Suzuki, L., J. P. Woessner, H. Uchida, H. Kuroiwa, Y. Yuasa *et al.*, 2000 A zygote-specific protein with hydroxyproline-rich glycoprotein domains and lectin-like domains involved in the assembly of the cell wall of *Chlamydomonas reinhardtii* (Chlorophyta). J. Phycol. **36:** 571–583.

Swanson, W. J., and V. D. Vaquier, 1998 Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. Science **281:** 710–712.

Thelen, J. J., G. M. Muzynski, J. A. Miernyk and D. D. Randall, 1998 Molecular analysis of two pyruvate dehydrogenase kinases from maize. J. Biol. Chem. **273:** 26618–26623.

Trickett, A. J., and R. K. Butlin, 1994 Recombination suppressors and the evolution of new species. Heredity **73:** 339–345.

Uchida, H., S. Kawano, N. Sato and T. Kuroiwa, 1993 Isolation and characterization of novel genes which are expressed during the very early stage of zygote formation in *Chlamydomonas reinhardtii*. Curr. Genet. **24:** 296–300.

Uchida, H., L. Suzuki, T. Anai, K. Doi, H. Takano *et al.*, 1999 A pair of invertedly repeated genes in *Chlamydomonas reinhardtii* encodes a zygote-specific protein whose expression is UV-sensitive. Curr. Genet. **36:** 234–240.

Umen, J. G., and U. W. Goodenough, 2001a Chloroplast DNA methylation and inheritance in *Chlamydomonas*. Genes Dev. **15:** 2585–2597.

Umen, J. G., and U. W. Goodenough, 2001b Control of cell division by a retinoblastoma protein homolog in *Chlamydomonas*. Genes Dev. **15:** 1652–1661.

Vanwinkle-Swift, K. P., R. Hoffman, L. Shi and S. Parker, 1994 A suppressor of a mating-type limited zygotic lethal allele also suppresses uniparental chloroplast gene transmission in *Chlamydomonas monoica*. Genetics **136:** 867–877.

Vanwinkle-Swift, K., K. Baron, A. McNamara, P. Minke, C. Burrascano *et al.*, 1998 The *Chlamydomonas* zygospore: mutant strains of *Chlamydomonas monoica* blocked in zygospore morpho-

genesis comprise 46 complementation groups. Genetics **148:** 131–137.

Von Heijne, G., J. Steppuhn and R. G. Herrmann, 1989 Domain structure of mitochondrial and chloroplast targeting peptides. Eur. J. Biochem. **180:** 535–545.

Weiss, R. L., D. A. Goodenough and U. W. Goodenough, 1977 Membrane differentiations at sites specialized for cell fusion. J. Cell Biol. **72:** 144–160.

Werner, E. B. E., W. R. Taylor and A. A. Holder, 1998 A *Plasmodium chabaudi* protein contains a repetitive region with a predicted spectrin-like structure. Mol. Biochem. Parasitol. **94:** 185–196.

Zou, J., Q. Qi, V. Katavic, E.-F. Marillia and D. C. Taylor, 1999 Effects of antisense repression of an *Arabidopsis thaliana* pyruvate dehydrogenase kinase cDNA on plant development. Plant Mol. Biol. **41:** 837–849.

Communicating editor: S. L. Allen