



HAL
open science

Megasatellites: a new class of large tandem repeats discovered in the pathogenic yeast *Candida glabrata*

Guy-Franck Richard, Agnès Thierry, Bernard Dujon

► To cite this version:

Guy-Franck Richard, Agnès Thierry, Bernard Dujon. Megasatellites: a new class of large tandem repeats discovered in the pathogenic yeast *Candida glabrata*. *Cellular and Molecular Life Sciences*, 2010, 67 (5), pp.671-6. 10.1007/s00018-009-0216-y . pasteur-01370699

HAL Id: pasteur-01370699

<https://pasteur.hal.science/pasteur-01370699>

Submitted on 23 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



MEGASATELLITES: A NEW CLASS OF LARGE TANDEM REPEATS DISCOVERED IN THE PATHOGENIC YEAST CANDIDA GLABRATA

Journal:	<i>Cellular and Molecular Life Sciences</i>
Manuscript ID:	Draft
Manuscript Type:	Visions and Reflections
Key Words:	Yeast, Minisatellite, Cell wall, Megasatellite, Genome, Replication, Recombination

1
2
3
4
5
6
7
8 **MEGASATELLITES: A NEW CLASS OF LARGE TANDEM REPEATS**
9
10
11 **DISCOVERED IN THE PATHOGENIC YEAST *CANDIDA GLABRATA***
12
13
14
15
16
17

18 **Agnès Thierry, Bernard Dujon and Guy-Franck Richard***
19

20
21
22
23 Institut Pasteur, Unité de Génétique Moléculaire des Levures; CNRS, URA2171; Université
24
25 Pierre et Marie Curie, UFR 927, 25 rue du Dr Roux, F-75015 Paris, France
26
27

28
29
30 Keywords: *Candida glabrata*, Minisatellite, Cell wall, Megasatellite, Genome
31
32
33

34
35 * Corresponding author
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Megasatellites are DNA tandem arrays made of large motifs, that were discovered in the yeast *Candida glabrata*. They are widespread in this species (40 copies) but are not found in any other hemiascomycete so far, raising the intriguing question of their very origin. They are found mainly in genes encoding cell wall products, suggesting that megasatellites were selected for a function linked to cell-cell adhesion or to pathogenicity. Their putative role in promoting genome rearrangements by interfering with DNA replication will also be discussed.

Megasatellites are a new class of DNA tandem repeats

In addition to its profound impact on evolutionary genomics or on our understanding of complex genetic networks, the systematic sequencing of whole eukaryotic genomes led to the discovery of new genetic elements. One of these discoveries was recently made in the genome of the opportunistic pathogen *Candida glabrata*. *C. glabrata* is a hemiascomycetous yeast, often involved in human candidiasis and bloodstream infections, particularly in immunocompromised patients [1,2]. *C. glabrata* is more resistant to fluconazole treatments than other pathogenic yeasts [3], and has become the second major causative agent of nosocomial infections due to yeast species. The *C. glabrata* genome of the reference strain (CBS138) was completely sequenced [4], and revealed that it was phylogenetically closer to *Saccharomyces cerevisiae* than to the other extensively studied pathogen *Candida albicans* [5]. We recently investigated the genome of *C. glabrata*, searching for minisatellites, a family of tandem DNA repeats whose motif size ranges from 9 nucleotides to usually less than 100 base pairs (reviewed in [6]). Besides the presence of numerous minisatellites, the *C. glabrata* genome also contains tandem repeats whose motif size is much longer, ranging from 135 to 417 nucleotides. We called this new family of large tandem repeats, megasatellites [7]. They

1
2
3 harbor two remarkable features: they are not found in any other sequenced living species
4
5 besides *C. glabrata* and *Kluyveromyces delphensis* (two *Saccharomycetaceae* yeasts of the
6
7 same clade [8]), and they are mainly found in genes proven, or suspected, to encode cell wall
8
9 proteins, raising the possibility that megasatellites could be directly involved in regulating cell
10
11 adhesion and pathogenicity. Altogether, 40 megasatellites were found in 33 genes in
12
13 *C. glabrata* and classified in two large families, called SFFIT and SHITT, following the
14
15 conservation of these five amino acids in each motif of the tandem repeat. Among these 40
16
17 megasatellites, 14 contain a SFFIT motif and 20 contain a SHITT motif, the number of motifs
18
19 in each tandem array ranging from 3 to 32 and covering from 405 to 9,600 DNA base pairs
20
21 (Figure 1). The remaining six motifs do not show obvious similarity with SFFIT and SHITT
22
23 families. Megasatellites are distributed on each of the 13 chromosomes but show some
24
25 preferential bias toward the subtelomeric regions, the right end of chromosome IX carrying
26
27 seven such elements within 65 kb, a density significantly higher than the genome average
28
29 (one megasatellite per 224 kb). Subtelomeric regions are highly flexible in
30
31 *Saccharomyces cerevisiae*, exhibiting a high level of inter-chromatid and inter-chromosome
32
33 recombination [9]. It is possible that *C. glabrata* subtelomeres share similar properties, and
34
35 that subtelomeric megasatellites recombine with each other, although this remains to be
36
37 demonstrated.
38
39
40
41
42
43
44
45
46
47

48 **Possible involvement of megasatellites in pathogenicity**

49
50 In *S. cerevisiae*, several genes involved in cell wall biogenesis contain minisatellites [10-12].
51
52 Some of them, called the *FLO* genes, play a direct role in flocculation and cellular adhesion.
53
54 It was shown that cell adhesion and flocculation is directly correlated to the length of the
55
56 minisatellite in *FLO1* [12], and that cell-cell adhesion leading to the formation of a biofilm at
57
58 the surface of sherry wine was directly dependent on the size of a minisatellite in *FLO11* [13].
59
60

1
2
3 In *C. glabrata* CBS138 strain, three *EPA* genes (functional homologues of the *FLO* genes,
4 [14-17]) contain megasatellites (*EPA2*, *EPA11* and *EPA13*), but three other *EPA* genes
5
6 contain simple minisatellites (*EPA1*, *EPA3* and *EPA15*), and three do not contain any kind of
7
8 tandem repeat (*EPA6*, *EPA7* and *EPA8*). In addition, 30 other genes that are not part of the
9
10 *EPA* family contain megasatellites. Some of the proteins encoded by these genes exhibit
11
12 signatures of cell-wall proteins but experimental evidence of their function or their
13
14 localization are lacking. Interestingly, in *S. cerevisiae*, three *FLO* genes (*FLO1*, *FLO5* and
15
16 *FLO9*) contain a 135-bp motif, tandemly repeated 7 to 13 times [11]. This Threonine-rich
17
18 motif shares no obvious similarity with any of the *C. glabrata* megasatellites. However, it has
19
20 the same size as the SHITT motif. It is therefore possible that 45 amino acids (135 bp) is the
21
22 optimal size for a tandem repeat in these cell-wall embedded proteins, and that the motif size
23
24 is therefore under strong selection, whereas the sequence itself is not necessarily conserved.
25
26
27
28
29
30

31 In budding yeast, telomeric regions are silenced by a multiprotein complex containing the *SIR*
32
33 genes, *RAP1*, *ESC1* and the Ku complex [18]. These genes are conserved in *C. glabrata*, to
34
35 the exception of *SIR1*, which is involved in silencing the silent mating-type loci, but not in
36
37 telomeric silencing [19]. The inactivation of *SIR3* and *RAP1* was shown to increase the level
38
39 of expression of several *EPA* genes, including the megasatellite-containing *EPA2* gene [15],
40
41 suggesting that the mechanism of subtelomeric silencing is probably similar in *C. glabrata*
42
43 and in *S. cerevisiae*. Subtelomeric megasatellite-containing genes are therefore probably also
44
45 silenced, although this remains to be shown. Therefore, at the present time, the possible role
46
47 played by megasatellites in *C. glabrata* pathogenicity is unclear, and needs to be clarified in
48
49
50
51
52
53 the future.
54
55
56
57
58
59
60

Megasatellites and genome rearrangements

Given the repeated nature of the large arrays formed by the megasatellites, one may wonder if they could behave like fragile sites and thus induce genome rearrangements. In humans, fragile sites are defined as chromatid constrictions or breaks visible on metaphasic chromosomes, when cells are grown in the presence of drugs that impair replication or DNA metabolism [6,20-22]. Although the precise molecular nature of all fragile sites is not known, some of them have been sequenced. *FRA3B* contains numerous transposons and LTRs found in direct and inverted orientations, *FRA10B* and *FRA16D* contain AT-rich minisatellites (42-bp and 33-bp motif size, respectively), *FRAXA*, *FRAXE*, *FRAXF*, *FRA11B* and *FRA16A* contain CGG trinucleotide repeats. Some of these fragile sites are associated to cancer. *FRA3B*, the most common fragile site in humans, often contains deletions in several gastrointestinal, colon, lung, breast and cervical cancers [23]. Loss of heterozygosity and a recurrent translocation were also observed at *FRA16D* in breast and prostate cancers and multiple myelomas [24]. Interestingly, chromosomal translocations and chromosome losses in *Candida albicans* are often associated to a large DNA tandem repeat, called the Major Repeat Sequence (MRS). It is a complex tandem repeat, found at nine different locations in the genome, and composed of a 2-kb motif tandemly repeated (RPS), itself including several tandem copies of smaller motifs (16 bp and 29 bp long). Most chromosome length polymorphisms in this yeast are due to size heterogeneity of the MRS [25]. When Muller and colleagues analyzed chromosomal translocations among different strains of *C. glabrata*, three major rearrangements, involving chromosomes IV, IX, XII and XIII, were found [26]. The three breakpoints corresponding to these three rearrangements were mapped and sequenced, but they are not located at the proximity of megasatellites, neither do they encompass any kind of repeated element. However, in the same study, it was shown that among 12 deletions, ranging in size from 130 bp to 12 kb, detected in the *C. glabrata* genome, two were located

1
2
3 within two megasatellites (one of them being the longest megasatellite of the genome). The
4
5 probability of this happening by chance being low, it suggests that megasatellites might be
6
7 involved in the mechanism leading to these two deletions. In a more recent study using a
8
9 larger number of probes, 11 reciprocal and non-reciprocal translocations involving 11 out of
10
11 the 13 *C. glabrata* chromosomes were found [27]. They also detected five segmental
12
13 duplications leading to the formation of new chromosomes. In one of these, a SFFIT
14
15 megasatellite is located less than 10 kb from a breakpoint, and could be involved in the
16
17 rearrangement, although this was not formally proven. In conclusion, observations made on
18
19 chromosomal plasticity in *C. glabrata* suggest that some of the rearrangements observed
20
21 might be triggered by the presence of a megasatellite. However, the majority of megasatellites
22
23 are not associated to chromosomal rearrangements and frequent rearrangements are observed
24
25 far from any megasatellite, showing that megasatellites are not systematically involved in
26
27 rearrangements. Large-scale studies of replication and recombination in *C. glabrata* are now
28
29 needed to understand the precise role of megasatellites in chromosomal replication and
30
31 instability.
32
33
34
35
36
37
38
39
40

41 **Evolution of megasatellites**

42
43 The last intriguing question concerning megasatellites relates to their mechanism(s) of
44
45 formation. One simple way to propagate megasatellites is to duplicate the gene(s) that contain
46
47 them. In *C. glabrata*, several megasatellites are found in paralogous gene families. The largest
48
49 of these families encompasses 11 paralogues, each containing a SHITT megasatellite and five
50
51 of them also containing a SFFIT repeat array (Rolland, Dujon and Richard, unpublished).
52
53 However, six megasatellites, containing either SHITT or SFFIT motifs, are found in genes
54
55 that are present in unique copies in the genome, raising the question of their very origin. If
56
57 point mutations followed by replication slippage may explain how smaller tandem repeats,
58
59
60

1
2
3 like microsatellites, are born, it is hard to imagine the same mechanism responsible for the *de*
4
5 *novo* creation of larger tandem arrays. Haber and Louis proposed that minisatellites are
6
7 formed by replication slippage between two short (5 bp) sequences flanking a 10-20 nt unique
8
9 sequence [28]. If most of the *S. cerevisiae* minisatellites are actually flanked by such short
10
11 motifs [11], it does not seem to be the case for *C. glabrata* megasatellites.

12
13
14
15 *In silico* comparisons of megasatellites with each other show that some motifs found in a
16
17 given gene are actually phylogenetically closer to motifs found in another gene, suggesting
18
19 that some kind of genetic transfer exists between megasatellites (Rolland, Dujon and Richard,
20
21 unpublished). This transfer may involve gene conversion, or alternatively one may imagine
22
23 that SHITT and SFFIT motifs are able to "jump" from one megasatellite to another one, using
24
25 mechanisms that may be related to transposition or retrotransposition (Figure 2). There is only
26
27 one full-size retrotransposon in the *C. glabrata* genome, a gypsy-like element (Tcg3, gene
28
29 name CAGLOG07183g, The Génolevures Consortium¹), two degenerate copies and two solo
30
31 LTRs (Cécile Neuvéglise, personal communication). Unless there is another source of reverse
32
33 transcriptase in this genome, it is difficult to hypothesize that retrotransposition is involved in
34
35 the spreading of megasatellites. In conclusion, the question of the origin of megasatellites is
36
37 still completely open, but experiments designed specifically to answer this question using
38
39 molecular tools available in this yeast species, should give some answers, and maybe explain
40
41 why SHITT and SFFIT megasatellites are so widespread in *C. glabrata*.

42 43 44 45 46 47 48 49 50 51 **Acknowledgements**

52
53 This work was supported by grant ANR-05-BLAN-0331 from the Agence Nationale de la
54
55 Recherche. B. D. is a member of the Institut Universitaire de France.

56
57
58
59
60
¹ <http://www.genolevures.org/>

- 1
- 2
- 3 [1] Bodey, G.P. et al. (2002) The epidemiology of *Candida glabrata* and *Candida albicans*
- 4 fungemia in immunocompromised patients with cancer *Am J Med* 112, 380-5.
- 5 [2] Raad, I. et al. (2004) Management of central venous catheters in patients with cancer
- 6 and candidemia *Clin Infect Dis* 38, 1119-27.
- 7 [3] Pfaller, M.A. and Diekema, D.J. (2004) Twelve years of fluconazole in clinical
- 8 practice: global trends in species distribution and fluconazole susceptibility of
- 9 bloodstream isolates *Clinical Microbiology and Infection* 10, 11-23.
- 10 [4] Dujon, B. et al. (2004) Genome evolution in yeasts *Nature* 430, 35-44.
- 11 [5] Dujon, B. (2006) Yeasts illustrate the molecular mechanisms of eukaryotic genome
- 12 evolution *Trends in Genetics* 22, 375-387.
- 13 [6] Richard, G.F., Kerrest, A. and Dujon, B. (2008) Comparative genomics and molecular
- 14 dynamics of DNA repeats in eukaryotes *Microbiol Mol Biol Rev* 72, 686-727.
- 15 [7] Thierry, A., Bouchier, C., Dujon, B. and Richard, G.-F. (2008) Megasatellites: a
- 16 peculiar class of giant minisatellites in genes involved in cell adhesion and
- 17 pathogenicity in *Candida glabrata* *Nucl. Acids. Res.* 36, 5970-5982.
- 18 [8] Kurtzman, C.P. (2003) Phylogenetic circumscription of *Saccharomyces*,
- 19 *Kluyveromyces* and other members of the *Saccharomycetaceae*, and the proposal of
- 20 the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and
- 21 *Zygorhizomyces* *FEMS Yeast Res* 4, 233-45.
- 22 [9] Louis, E.J., Naumova, E.S., Lee, A., Naumov, G. and Haber, J.E. (1994) The
- 23 chromosome end in yeast: its mosaic nature and influence on recombinational
- 24 dynamics *Genetics* 136, 789-802.
- 25 [10] Bowen, S., Roberts, C. and Wheals, A.E. (2005) Patterns of polymorphism and
- 26 divergence in stress-related yeast proteins *Yeast* 22, 659-668.
- 27 [11] Richard, G.-F. and Dujon, B. (2006) Molecular evolution of minisatellites in
- 28 hemiascomycetous yeasts *Mol Biol Evol* 23, 189-202.
- 29 [12] Verstrepen, K.J., Jansen, A., Lewitter, F. and Fink, G.R. (2005) Intragenic tandem
- 30 repeats generate functional variability *Nature Genetics* 37, 986-990.
- 31 [13] Fidalgo, M., Barrales, R.R., Ibeas, J.I. and Jimenez, J. (2006) Adaptive evolution by
- 32 mutations in the *FLO11* gene *Proc Natl Acad Sci U S A* 103, 11228-11233.
- 33 [14] Castano, I., Pan, S.-J., Zupancic, M., Hennequin, C., Dujon, B. and Cormack, B.P.
- 34 (2005) Telomere length control and transcriptional regulation of subtelomeric
- 35 adhesins in *Candida glabrata* *Molecular Microbiology* 55, 1246-1258.
- 36 [15] De Las Penas, A., Pan, S.J., Castano, I., Alder, J., Cregg, R. and Cormack, B.P. (2003)
- 37 Virulence-related surface glycoproteins in the yeast pathogen *Candida glabrata* are
- 38 encoded in subtelomeric clusters and subject to *RAP1*- and *SIR*-dependent
- 39 transcriptional silencing *Genes Dev* 17, 2245-58.
- 40 [16] Frieman, M.B., McCaffery, J.M. and Cormack, B.P. (2002) Modular domain structure
- 41 in the *Candida glabrata* adhesin *Epa1p*, a β 1,6 glucan-cross-linked cell wall protein
- 42 *Molecular Microbiology* 46, 479-492.
- 43 [17] Zupancic, M., Frieman, M.B., Smith, D., Alvarez, R.A., Cummings, R.D. and
- 44 Cormack, B.P. (2008) Glycan microarray analysis of *Candida glabrata* adhesin ligand
- 45 specificity *Molecular Microbiology* 68, 547-559.
- 46 [18] Taddei, A., Hediger, F. and Gasser, S.M. (2004) The function of nuclear architecture:
- 47 a genetic approach *Annual Review of Genetics* 38, 305-345.
- 48 [19] Fabre, E., Muller, H., Therizols, P., Lafontaine, I., Dujon, B. and Fairhead, C. (2005)
- 49 Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and
- 50 subtelomeres *Mol Biol Evol* 22, 856-873.
- 51 [20] Debacker, K. and Kooy, R.F. (2007) Fragile sites and human disease *Hum Mol Genet*
- 52 16 Spec No. 2, R150-8.
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

- 1
2
3 [21] Glover, T.W., Arlt, M.F., Casper, A.M. and Durkin, S.G. (2005) Mechanisms of
4 common fragile site instability Hum Mol Genet 14 Spec No. 2, R197-205.
5 [22] Sutherland, G.R., Baker, E. and Richards, R.I. (1998) Fragile sites still breaking TIG
6 14, 501-506.
7 [23] Durkin, S.G., Ragland, R.L., Arlt, M.F., Mülle, J.G., Warren, S.T. and Glover, T.W.
8 (2008) Replication stress induces tumor-like microdeletions in FHIT/FRA3B Proc
9 Natl Acad Sci U S A 105, 246-51.
10 [24] Popescu, N.C. (2003) Genetic alterations in cancer as a result of breakage at fragile
11 sites Cancer Letters 192, 1-17.
12 [25] Magee, P.T. (2007) in: Candida comparative and functional genomics, pp. 7-26
13 (d'Enfert, C. and Hube, B., Eds.) Caister Academic Press.
14 [26] Muller, H. et al. (2008) Genomic polymorphism in the population of *Candida*
15 *glabrata*: gene copy-number variation and chromosomal translocations Fungal
16 Genetics and Biology In press.
17 [27] Polakova, S., Blume, C., Zarate, J.A., Mentel, M., Jorck-Ramberg, D., Stenderup, J.
18 and Piskur, J. (2009) Formation of new chromosomes as a virulence mechanism in
19 yeast *Candida glabrata* Proc Natl Acad Sci U S A 106, 2688-93.
20 [28] Haber, J.E. and Louis, E.J. (1998) Minisatellite origins in yeast and humans Genomics
21 48, 132-135.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure legends

Figure 1: Schematic representation of *C. glabrata* SHITT and SFFIT megasatellites.

Gene names are indicated to the left (<http://www.genolevures.org>). Vertical lines near gene names indicate paralogous gene families. The region of the gene containing the megasatellite is represented by colored boxes. Orange: SHITT motifs. Green: SFFIT motifs. The number in boxes corresponds to the number of tandemly repeated motifs (no number indicates the presence of only one motif). Purple: degenerate SFFIT motifs. The degenerate SFFIT motifs found within CAGL0J11924g and CAGL0J05170g are not identical, although they both probably come from a SFFIT motif. Light grey: sequence of variable length, that is sometimes tandemly repeated and found interspersed within some of the SHITT and SFFIT megasatellites. Dark grey: Glycine- and Serine-rich motifs, of variable length, found only in CAGL0L00227g and CAGL0J01774g. Their size (in amino-acids) is indicated when too long to be drawn to scale. Otherwise, boxes are drawn to scale. Numbers shown before and after boxes represent the number of amino-acids before and after the repeated motifs. Given the lower sequence coverage and unprecise assembly of subtelomeric regions, it is possible that the number of motifs shown for the 11 subtelomeric megasatellite-containing genes is different from what is represented here (11 genes, from CAGL0A04873g to CAGL0L00227g). (1) The subtelomeric sequence is interrupted within a megasatellite.

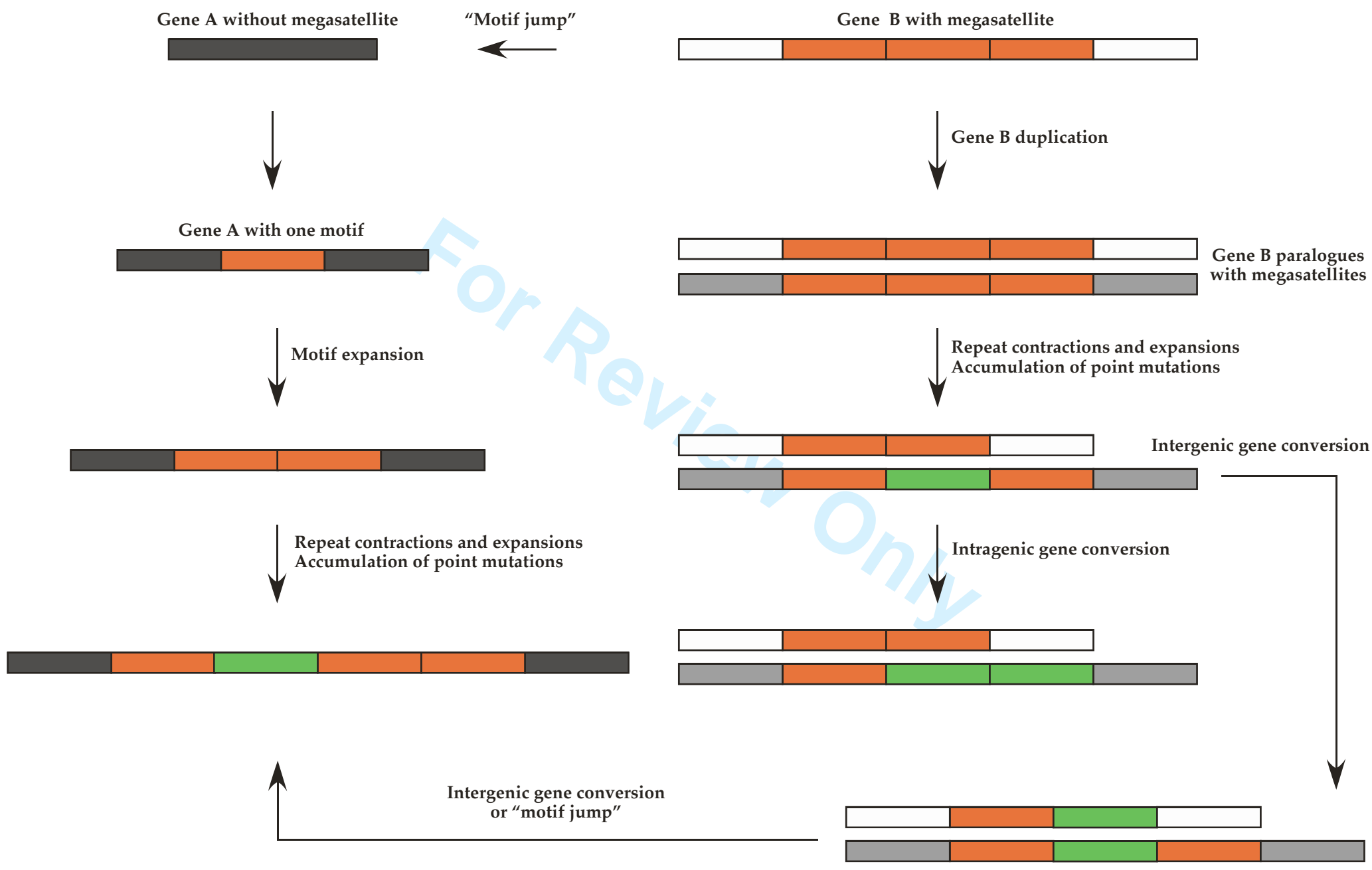
Figure 2: Different mechanisms can lead to megasatellite spreading in the *C. glabrata* genome.

Left: gene A with no megasatellite may acquire a motif by retrotransposition or another mechanism, followed by expansion of the motif into a megasatellite. Right: a gene already containing a megasatellite may duplicate itself, leading to the formation of two paralogues, each of them containing an identical megasatellite. Contractions and expansions may now

1
2
3 occur, independently in both tandem repeats, and point mutations may accumulate in one or
4
5 several motifs, leading to slightly different motif sequences (in green). These new motifs may
6
7 propagate (or disappear) by intergenic (or intragenic) gene conversion. New motifs may also
8
9 propagate by "jumping" into a megasatellite encoded by a non-paralogous gene (bottom).
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47



Thierry *et al.*
Figure 2