Figure S8

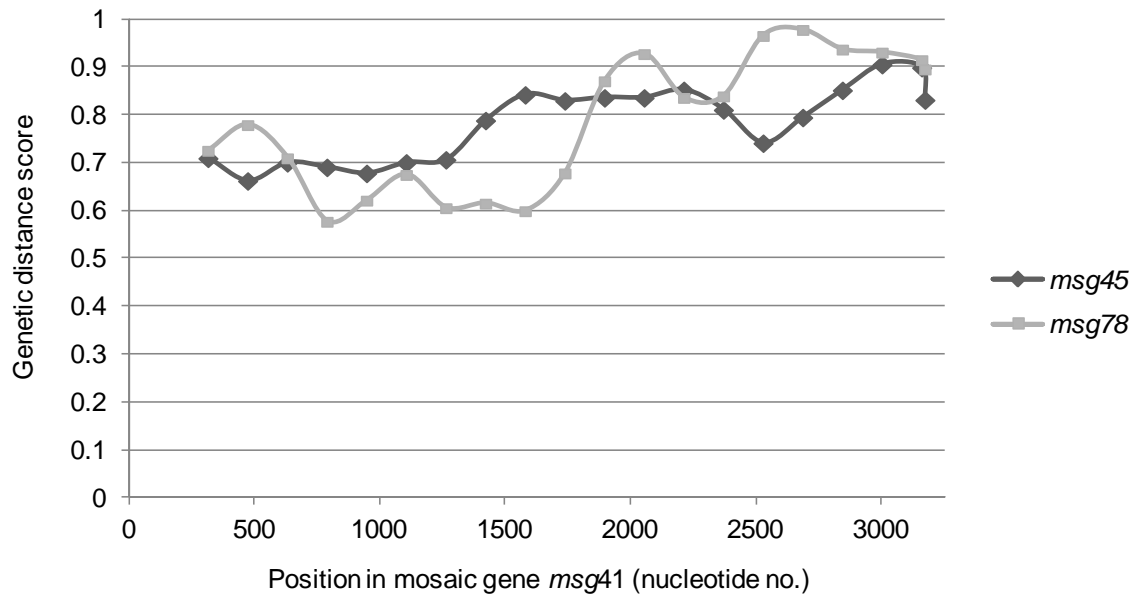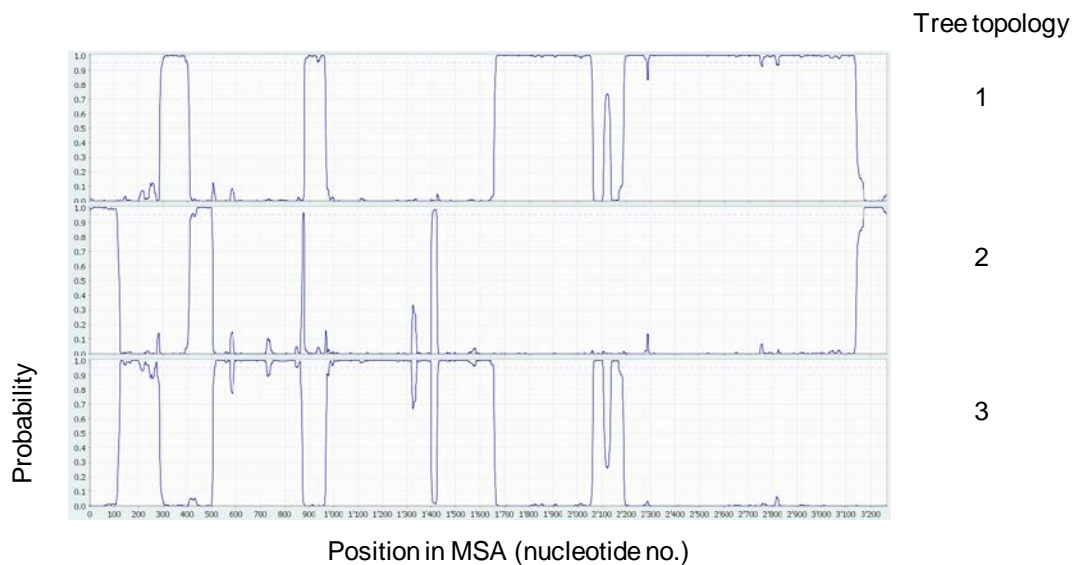**Detection of potential mosaic genes sharing an identical fragment with one putative parent gene.**

a, Mosaic gene *msg*41. **b,** Mosaic gene *msg*104. **c**, Mosaic gene *msg*79.

a1

Genetic distance score (y-axis), Position in mosaic gene *msg*41 (nucleotide no.) (x-axis)

— msg45
— msg78

a1, the set of 11 full length *msg*-I genes was analysed using the Recombination Analysis Tool. The predicted recombination sites are at positions ca. 600, 1900, and 2200. The Bellerophon method detected a potential recombination event at position 2200.

a2

Tree topology

1

2

3

Probability (y-axis), Position in MSA (nucleotide no.) (x-axis)

a2, analysis of the mosaic gene *msg*41 with its putative parent genes together with the randomly chosen *msg*56 of the same family using TOPALi based on the Hidden Markov Model. Recombination events are also detected at positions ca. 600, 1900, and 2200, but several other recombination events are predicted.
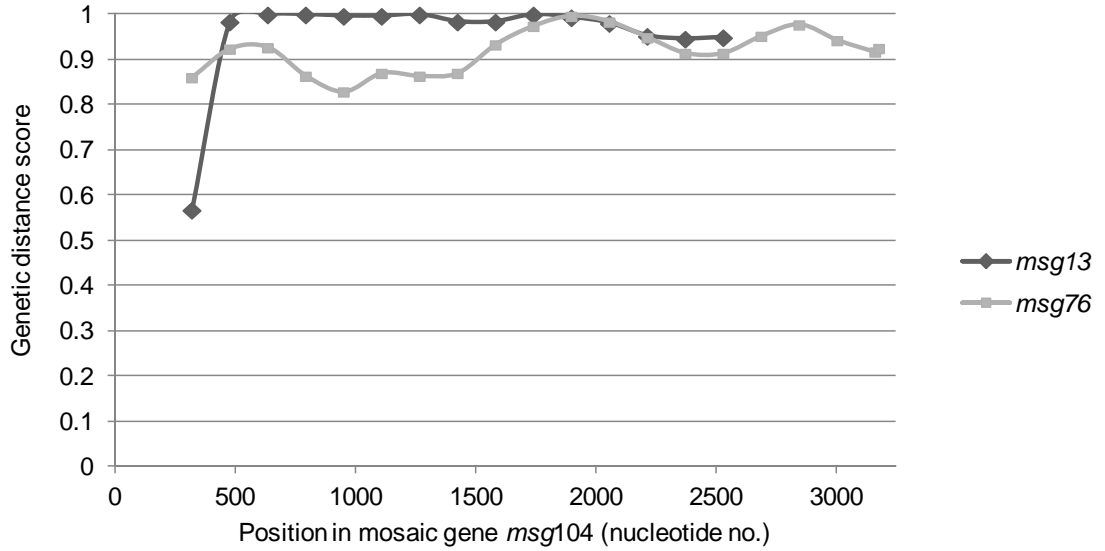
a3

```
msg41   TGATGGTAAAAGACGTACAAGATAGGTGCAAAGTATTTAAGGAAAATATA  1917
msg78   TGATGGTAAAAGACGTGCAAGATAGGTGTAAAGTATTCAAAGAAAATATA  1959
        *************** ********** ******** ** *********

msg41   CAAGTATCAGAGATCGTTGAGTTTCTTAAAAATAACACGAATAACATAAC  1967
msg78   AAAGCTTCAGAGATTGTTGATTTTCTTAAAAATAACACGAATAACATAAC  2009
         ***  ******** ***** ****************************

msg41   AACATTAGGAAATGTTTGCCCATTTTGGGACCCATATTGTGACAAATTTT  2017
msg78   AACACTGGAAAGAAACTGTCCCTCCTGGCATACGTACTGCAATAGATTTT  2059
        **** * **    ** ** *  *** *  * ** **  * * *****

msg41   CGCCTAATTGTCCAGATCTTACGAAAAAAAAATACTCTTTGTACAAAGATC  2067
msg78   CACCTAATTGTCCAGATCTTACGAAAGAGGATAG---TTGTACAAAGTC   2106
        * ******************* *  * ***   ********* **

msg41   AAAAGACCATTGTAAGCCATTTTATGAAAGAAAGGCCTTAGAAGATGCTCT  2117
msg78   AAGAAACATTGTGAGCCGTTCTATAAAAGAAAGGCCTTGGAAGATGCTCT  2156
        **  * ****** **** **  ** *************  **********

msg41   CAAAGTAAAGCTTCGAGGGAAAGCTGAGTGATGAAAACAAATGTACTACAG  2167
msg78   CAAAGTAGAGCTTCGAGGGAAAGCTGAGTGATAAAAACAAATGTACTACAG  2206
        ******* ************************ *****************

msg41   AATTAAAG-GATATTGTACAA-AGCGAA-AACGTAAATAATGCGTCAATC  2214
msg78   AATTAAAGGGATATTGTACAATAGCGAAAAACGTAAATAATGCGTCAATC  2256
        ******** ***********  ****** **********************

msg41   AGTGGTTTATGCAAAGATAGCACTCAAGGTAATAATAAAAAGCCTGACGA  2264
msg78   AATGGTTTATGCAAAGATAACACTCAAGGTAATAATAAAAAGCCTGACGA  2306
        * ***************** ******************************

msg41   CAAAGTGGTAGAAGAGCTTTGTAAGAAATTGATGGAAGAAATAAAAGAGC  2314
msg78   CAAAGTGGTAGAAGAGCTTTGTAAGAAATTGATGGAAGAAATAAAAGAGC  2356
        *************************************************

msg41   AATGTGAGACATTACCAGCAGAATTAACGGAGCTGGAAAAAAGCCTAGAA  2364
msg78   AATGTGAGACATTACCAGCAGAATTAACGGAGCTGGAAAAAAGCCTAGAA  2406
        *************************************************

msg41   AAAGATGTTAAAACATATAAAGAACTTAAGAAAGAGGCAAAGAAAGCAAT  2414
msg78   AAAGATGTTAAAACATATAAAGAACTTAAGAAAGAGGCAAAGAAAGCAAT  2456
        *************************************************

msg41   GAACAAGTCTAATCTTGTTTTATCACTCGTTAAGAAAAACGAAAGTAATA  2464
msg78   GAACAAGTCTAATCTTGTTTTATCACTCGTTAAGAAAAACGAAAGTAATA  2506
        *************************************************

msg41   CATCAAAAAATAATAGAAATAATAGCAAAAACAAGGATAAGAATGCCGTT  2514
msg78   CATCAAAAAATAATAGAAATAATAGCAAAAACAAGGATAAGAATGCCGTT  2556
        *************************************************

msg41   TCAAATGGACTTCAAGATACCACAGAACATGTGAAAATACTACGGAG-GG  2563
msg78   TCAAACGGATTTCAAGATATCACCGAACATGTGAAAATACTACGGAGAGG  2606
        ***** *** ********* *** ******************** **
```

```
msg41   AGTTAAGGAGGCACTTG--ACAGAATCTGAAGCCAAGGCATTTGATTTAG  2611
msg78   AGTTAAGGAGGCACTTGTAACAGAATCTGAAGCCAAGGCATTTGACCTAG  2656
        ****************  *********************** ***

msg41   CTGCAGAAGTATTTGGAAGATATGTAGACTTGAAAGAAAGATGTGAGAAA  2661
msg78   CAGCAGATGTGTTTGGAAGATATGTAGATTTGAAGGAAAGATGTGAGAAA  2706
        * ***** ** ***************** ***** **************

msg41   TTGACCTCAGGTTGCGGGATTAAAGACGATTGTAAAGATCTAGAAAATGT  2711
msg78   TTAAAATCGGATTGCGGGATTAAAGACGATTGTAAAGATCTAGAAAATGT  2756
        ** *  ** * *********************************

msg41   ATGCAAAAAGATCGGGAAAACATGTAGCGATCTGAAGCCTCTGGAGGTGA  2761
msg78   ATGCAAAAAGATCGAGAAAACATGTAGCGATCTGAAGCCTCTGGAGGTGA  2806
        ************* ***********************************

msg41   AGTCGCACGAAATAGTCACAGAAAGCACAACAACGACCACAACGACAACA  2811
msg78   AGCCACACGAAATAGTTACAGAAAGCACAACGACCACGACGACAACA    2856
        ** * ********** ************** ******** *********

msg41   ACAACCGTTACCGATCCGAAGGCAACAGAATGCAAATCCTTACAGACAAC  2861
msg78   ACAACCGTTACCGATCCGAAGGCAACAGAATGCAAATCCTTACAGACAAC  2906
        *************************************************

msg41   AGACACATGGGTTACACAGACATCGACACACACGAGTACATCTACCATCA  2911
msg78   AGATACATGGGTTACACAGACATCGACACACACAAGCACGTCTACCATCA  2956
        *** ***************************** ** ** *********

msg41   CATCTACGATTACATCAAAAATAACACTCACATCAACGAGGCGGTGCAAA  2961
msg78   CATCTACGATTACATCAAAAATAACACTCACATCAACGAGGCGGTGCAAA  3006
        *************************************************

msg41   CCAACCAAGTGTACGACAGGGGATGATGCAGAAGACGTGAAGCCGAGTGA  3011
msg78   CCAACCAAGTGTACGACAGGGGATGATGCAGAGGACGTGAAACCGAGTA-  3055
        ******************************** ******** *****

msg41   GGGATTGAAGATGAGTGGGTGGAGCGTGATGAGGGGGGTGATATTAGCAA  3061
msg78   -GGCTTGAGGGTGAGCGGGTGGAATGTGATGAGGG----GATAGTAGCAA  3100
        **   *** **** ********   ********      **** ******

msg41   TGATGATTTCGTTCATGATTTAG  3084
msg78   TGGTTATTTCGTTCATGATTTAG  3123
        ** * ******************
```

a3, alignment of the mosaic gene *msg*41 and its parent *msg*78. The two genes share the 285 bps 100% identical from position 2234 to 2519 in *msg*41 which are underlined, after the predicted recombination site at position ca. 2200.

b1, the set of 11 full length *msg*-II genes was analysed using the Recombination Analysis Tool together with *msg*13 of the same family which is truncated by the end of the contig at position 2442. The predicted recombination site is at position ca. 400. The Bellerophon method did not identify this mosaic gene.

b2, analysis of the mosaic gene *msg*104 with its putative parent genes together with the randomly chosen *msg*62 of the same family using TOPALi based on the Hidden Markov Model. A recombination event is detected at position 200, *i.e.* close to that detected with the screening method at ca. 400.

b3

```
msg104  ATGAAAGCATTTGCATTAGCCAGCTTTCTTGGCATAGCGTATGCTTTTTC 50      msg104  AAAGAAATTATATTAATAGAAAAAGTTGGAAAAGAAAAAATATTCAAAGA 950
msg13   ATAATA--------CTAGA-----------------------------  11      msg13   AAAGAAATTATATTAATAGAAAAAGTTGGAAAAGAAAAAATATTCAAAGA 839
        ** * *         ***                                              **************************************************

msg104  AAAAAACATTGAATCTATTTATAAAGAATCTTTAAATCCATTACAAAATC 100     msg104  CGAAGTTGGAAAACCAGGGATAAAAGATACAATTGATCTTTTAGTATTAT 1000
msg13   -----AC-----------TTGTAAA-------------ATTA------  24      msg13   CGAAGTTGGAAAACCAGGGATAAAAGATACAATTGATCTTTTAGTATTAT 889
         **              ** ****              ****                      **************************************************

msg104  CACCACAAAATCCACCACAAAATCCGTCACAAAATCCATTATTGTATATA 150     msg104  TATCAAATAATTATTTAGATGATTGCAAAAATTATATTGAAAAGTGTCAT 1050
msg13   ----A------------CA-----------AA-----TTATT-TATT-T  39      msg13   TATCAAATAATTATTTAGATGATTGCAAAAATTATATTGAAAAGTGTCAT 939
         *           **          **  ***** ***                          **************************************************

msg104  ACTGAAGGAAATATTTTTGCTTTAATTCTAAAAAATGCAATGAATACTGA 200     msg104  AAACTTTGCAGCCTTTTACCTCAATTAGAGGATTTATATGACAGTGCTAA 1100
msg13   AGTGAAGGAAATATTTTTGCTTTAATTCTAAAAAATGCAATGAATACTGA  89      msg13   AAACTTTGCAGCCTTTTACCTCAATTAGAGGATTTATATGACAGTGCTAA 989
         *  *********************************************               **************************************************

msg104  ATGTAAAGCAAGGCTAAGAAAATACTGTGAAAATTTAAGAAATATGACTC 250     msg104  GGAAAAAATGAATAAAAGCAAAGAAGAAGTATGTGATACCTTAAAAGAAA 1150
msg13   ATGTAAAGCAAGGCTAAGAAAATACTGTGAAAATTTAAGAAATATGACTC 139     msg13   GGAAAAAATGAATAAAAGCAAAGAAGAAGTATGTGATACCTTAAAAGAAA 1039
        **************************************************              **************************************************

msg104  AAAATGCTAAAAAGCTCCTTCAAAGCATTAGAAGAACTATGCCAAGAAACG 300    msg104  AATTAAAGCCTAAATGTAGGTTTTTTAAATCAAAATTATATGATTTATCA 1200
msg13   AAAATGCTAAAAAGCTCCTTCAAAGCATTAGAAGAACTATGCCAAGAAACG 189    msg13   AATTAAAGCCTAAATGTAGGTTTTTTAAATCAAAATTATATGATTTATCA 1089
        **************************************************              **************************************************

msg104  AAACTAGATGAAAAGTGCAATGATTTAAAAGAAAATATTATGAGAAGATG 350     msg104  CTATCAAATACTAAAAAAGACAATGAAGATGCTACTCTTATAAAATGGAC 1250
msg13   AAACTAGATGAAAAGTGCAATGATTTAAAAGAAAATATTATGAGAAGATG 239     msg13   CTATCAAATACTAAAAAAGACAATGAAGATGCTACTCTTATAAAATGGAC 1139
        **************************************************              **************************************************

msg104  TGCTGCACTTGAACGAAGTTTGCAAGATATCCAGACTAAACCATCTCCAA 400     msg104  CAAGCAATCTACTGAGTTTAATGAAAAGCTTTGTATAAATTTAGAATCAA 1300
msg13   TGCTGCACTTGAACGAAGTTTGCAAGATATCCAGACTAAACCATCTCCAA 289     msg13   CAAGCAATCTACTGAGTTTAATGAAAAGCTTTGTATAAATTTAGAATCAA 1189
        **************************************************              **************************************************

msg104  GTACTTTTGAATGTCTGTATCAAAATGAATGTATGCGTATAGAAGAAGCA 450     msg104  AATGTTTTTATTTAAGGAAATCTTGTAATAATGCGGGTATTAAAATGTCT 1350
msg13   GTACTTTTGAATGTCTGTATCAAAATGAATGTATGCGTATAGAAGAAGCA 339     msg13   AATGTTTTTATTTAAGGAAACCATGTAGTGATGAGGGTATTAAAATGTCT 1239
        **************************************************              ******************** * **** * *** ***************

msg104  TGGTTCAATAAAAGTTAATGAAAAATGCAATTATTTAAGAGCATTTTGTAG 500    msg104  AATGCATGTATCAATGTAGAATCAACATGTTTAAAAACACGACTTTTCAG 1400
msg13   TGGTTCAATAAAAGTTAATGAAAAATGCAATTATTTAAGAGCATTTTGTAG 389    msg13   AATGCATGTATCAATGTAGAATCAACATGTTTAAAAACACGACTTTTCAG 1289
        **************************************************              **************************************************

msg104  AGAAAAAAGACGAGATGATTTGAAAACAGAATTTCTTTTAAGAGCCTTAT 550     msg104  AAGAGAATATCAACTATTTCAGAGCACATTAAAAGGGAAACTGCATAATT 1450
msg13   AGAAAAAAGACGAGATGATTTGAAAACAGAATTTCTTTTAAGAGCCTTAT 439     msg13   AAGAGAATATCAACTATTTCAGAGCACATTAAAAGGGAAACTGCATAATT 1339
        **************************************************              **************************************************

msg104  CTGGAAATTTAAAAACTCAAGAAGAGTGTGAGAAAATAATCGACAAAAAG 600     msg104  TAACAAATAATTCGCTTAAAACGTGCATAGATGAATTATGGATTCTATGT 1500
msg13   CTGGAAATTTAAAAACTCAAGAAGAGTGTGAGAAAATAATCGACAAAAAG 489     msg13   TAACAAATAATTCGCTTAAAACGTGCATAGATGAATTATGGATTCTATGT 1389
        **************************************************              **************************************************

msg104  TGTCTTGCGTTCATGGGAGAAAGCGATGAACTGATGAAGTTTTGTTTAAT 650     msg104  GAGAAAATAATTAGCAGTAATAATCCTATATTAATAGATCTTTGTCTACA 1550
msg13   TGTCTTGCGTTCATGGGAGAAAGCGATGAACTGATGAAGTTTTGTTTAAT 539     msg13   GAGAAAATAATTAGCAGTAATAATCCTATATTAATAGATCTTTGTCTACA 1439
        **************************************************              **************************************************

msg104  CCCCCTTAAATAGATGTAACGAACTCGTGAAATTAATGGAGACAAAATGTA 700
msg13   CCCCCTTAAATAGATGTAACGAACTCGTGAAATTAATGGAGACAAAATGTA 589
        **************************************************

msg104  ATAGTTTAGAATTTGAGGCAAAACTTTTTATCGAAAAGACTAAAATATCA 750
msg13   ATAGTTTAGAATTTGAGGCAAAACTTTTTTATCGAAAAGACTAAAATATCA 639
        **************************************************

msg104  AAAGAAGAATGCATTTCATTATTTGAGAAATGCTATTTTCATAAATCAAA 800
msg13   AAAGAAGAATGCATTTCATTATTTGAGAAATGCTATTTTCATAAATCAAA 689
        **************************************************

msg104  TTGCAATAACACTCTTGATAATAAATGTAGAGAAATAGAAAAAAAGTGTG 850
msg13   TTGCAATAACACTCTTGATAATGAATGTAGAGAAATAGAAAAAAAGTGTG 739
        **********************  **************************

msg104  AAAATGAAGTAGAATACAAATATTTTTCTTTACCATTTAATCCAGTAGGA 900
msg13   AAAATGAAGTAGAATACAAATATTTTTCTTTACCATTTAATCCAGTAGGA 789
        **************************************************
```

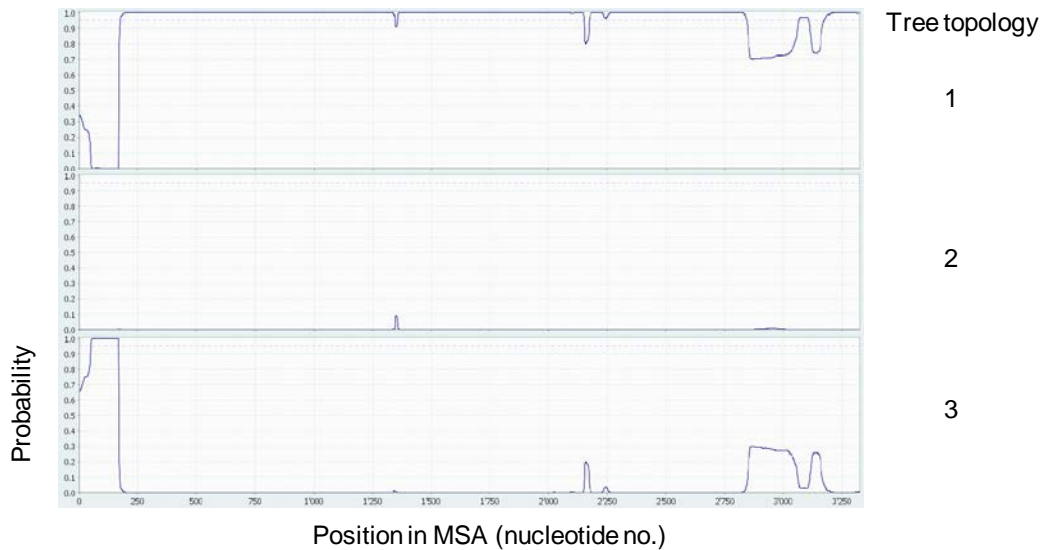**b3**, alignment of the mosaic gene *msg*104 and its parent *msg*13. The two genes share 1167 bps 100% identical from position 153 to 1320 in *msg*104 which are underlined, after the potential recombination site at position ca. 200.

c

```
msg79   ATGCTAAGGAATTAGAGGAAATTTTGTTAGAAGAAAAAGTAGAAAAATTTA 1797    msg79   AGGAATGTAAATTTAAAAACAGATGCAACTCAGAAAAATCAATCAAGGTA 2597
msg7    ATGCGAAAGAGTTAGAAGATATTTTATTAGAGGAAAAAGCAGAGAATTTA 1797    msg7    AGAAATGTAAATTTAAAAACAGATGTAACTCTGAAAAATCAATCAAGGTA 2597
        **** ** ** ***** ** ***** ***** ******* *** ******                ** ************************ ***** ******************
msg79   GGCAATTTAGATACATGTATAAAAAAGTTTCAGAAAAATGTAATAAATG 1847    msg79   TTTGGAACAAAAGGAGACAAAAGTAAATATTACAGAAAAAGAAGTTGAAG 2647
msg7    GGCAATTTAGATATATGTATAGAAAGAGTTTCAGAAAAATGTAATAAATG 1847    msg7    TTTGGAACAAAAGGAGACAAAAGTGAATATTACAGAAAAAGAAGTTGAAG 2647
        ************* ******* *** *************************                ************************ **************************
msg79   GTCTAAAAAAAAGAGAACAAGATTTATTTTTTCGTGTATACAATTAGTCA 1897    msg79   CGTTTGATGCAGCAGCAGAAGCATTAAAGGTTTATACAGAAGTCAAAGCA 2697
msg7    GTCTAAAAAAAAGAGAACAAGATTTATTTTTTCGTGTATACAATTAGTCA 1897    msg7    CGTTTGATGCAGCAGCAGAAGCATTAAAGGTTTATACAGAAGTCAAAGCA 2697
        **************************************************                **************************************************
msg79   CTACTTGTCAAATCATTACTAGAGACATTAAATCTAAATGTTCTATATTA 1947    msg79   GAGTGCAAAGGTCTGCAGTTAGAATGTGAATTTAAGGAAGATTGTTCTGA 2747
msg7    CTACTTGTCAAATCATTATTAGAGATATTAAATCTAAATGTTCTATATTA 1947    msg7    GAGTGCAAAGGTCTGCAATTAGAATGTGAATTTAAGGAAGATTGTTCTGA 2747
        ****************** ****** ************************                ***************** ******************************
msg79   GAAAGGAACATAGATATTGAAGACGTTCTAGATCAAATGAAGAATGATAA 1997    msg79   ATACAAAGATGTATGCAAGAAAATTGAAGATGCATGCAATAAATTAAAGT 2797
msg7    GAAAGGAACATAGATATTGAAGACGTTCTAGATCAAGTGAAGAATGATAA 1997    msg7    ATACAAAGATGTATGCAAGAAAATTGAAGATGCATGCAATAAATTAAAGT 2797
        ************************************ *************                **************************************************
msg79   TGCAGACATAAAGGGCTCTACATGTGATTTATGGGAACCCTATTGTGATA 2047    msg79   CACTGGAAATCAAGTCTTCAGAAACAAAAACAATTAATCAAACTATAAAG 2847
msg7    TGCAGACATAAAGGGCCCTACATGTGATCTATGGGAACCCTATTGTAATA 2047    msg7    CACTGGAAATCAAGTCTTCAGAAACAAAAACAATTAATCAAACTATAAAG 2847
        **************** ***********  ***************** ***                **************************************************
msg79   AGTTTATGTTGAGCTGCGAAATACTTGTACAAAATAATGGAAAGAATGGA 2097    msg79   ACCATCATTACAAAAACCGAGACAAATACAACACAAAAAACACTGACGAC 2897
msg7    AGTTTATGTTGAGCTGCGAAATACTTGTACAAAATAATGGAAAGAATGGA 2097    msg7    ACCATCATTACAAAAACCGAGACAAATACAACACAAAAAACACTGACGAC 2897
        **************************************************                **************************************************
msg79   AAGTGCAAAGAACTAAAGGAGAGCTGTGAACCCTATCGTAGAATACAAGA 2147    msg79   AGGAGAACAATGTATGTCGATTTCTACAACAGACAAATGGATTACGCGTA 2947
msg7    AAGTGCAAAGAACTAAAGGAGAGCTGTGAACCCTATCGTAGAATACAAGA 2147    msg7    AGGAGAACAATGTATGTCGATTTCTACAACAGACAAATGGATTACGCGTA 2947
        **************************************************                **************************************************
msg79   ACAAGAAATGAAACTTATGTATGAATTAAGAGGAGGTCTCAATAGCGAAA 2197    msg79   CATCAACACACACACACATCTACACAAACATCCGTACTAACATTAACT 2997
msg7    ACAAGAAATGAAACTTATGTATGAATTAAGAGGAGGTCTCAATAGCGAAA 2197    msg7    CATCAACACACACACACACATCTACACAAACATCCGTACTAACACTAACT 2997
        **************************************************                ************************************** *** *****
msg79   ATAAATGTAAATCAACCCTTAATGAACACTGTTTGCATTGGGATAAAACG 2247    msg79   GTAACATTAACATCAACGAAGGAATGTAAACCAATGAAGTGCACTACCGG 3047
msg7    ATAAATGTAAATCAACCCTTAATGAACACTGTTTGCATTGGGATAAAACG 2247    msg7    GTAACATTGACATCAACGAAGGGATGTAAACCAGTTAAATGCACTATTGG 3047
        **************************************************                ******** ************* ********** * ** ******* **
msg79   AAAAAATGATACATTCAAGAATTTCTGCAACAATAACACTGATACCAAAAA 2297    msg79   GAGTGAAGAGGAAGCAGGAGATGTAAAATCAAGTAAAGGGTTAAGGATGA 3097
msg7    AAAAAATGATACATTCAAGAATTTCTGCAACAATAACACTGATACCAAAAA 2297    msg7    AAGTGGAGAAGAAGCAGGAGATGTAAAACCAAGTAAAGGGTTAAGGATGA 3097
        *************************************************                 **** *** ******* ********** * ** ******* **
msg79   TAATACAACTAAAAATGAACTGTGCAAGAAACTACTAAAACATGTAAAGG 2347    msg79   ATGGGTGGAGCCTAATAAAAAGAGTAATATTAATAATGATTATTTCAACT 3147
msg7    TAATACAACTAAAAATGAACTGTGCAAGAAATTACTAAAATATGTAAAGG 2347    msg7    ATGGATGGGGTTTGATAAAAGGAGTAATATTAACAATGATTATTTCAACT 3147
        ******************************* ******** *********                **** *** * *  * ****** *********** ****************
msg79   AAAGATGTACAAAATTATTGGCAAAATTAAATGGCATGGCAACGGAAATA 2397    msg79   ATGATTTAA 3156
msg7    AAAGATGTACAAAATTATTGGCAAAATTAAATGGCATGGCAAGGGAAATA 2397    msg7    ATGATTTAA 3156
        ****************************************** *******                *********
msg79   GAAGAAAATGTAAAGATTGTTGAAAAATTGAACGAAGCAGCGAAAAAAGC 2447
msg7    GAAGAAAATGTAAAGATTGTTGAAAAATTGAACGAAGCAGCGAAAAAAGC 2447
        **************************************************
msg79   ACTAAAGAATACAAATCTTATTTTAACTTCAAGCAAACAAAAAACAGATT 2497
msg7    ACTAAAGAATACAAATCTTATTTTAACTTCAAGCAAACAAAAAACAGATT 2497
        **************************************************
msg79   CTAACATAAATAACGCAACATTAATTCTAGCTTATAATGCAAATGCTGAC 2547
msg7    CTAACATAAATAACGCAACATTAATTCTAGCTTATAATGCAAATGCTGAC 2547
        **************************************************
```

c, alignment of the region of the mosaic gene *msg*79 and its parent *msg*7 covering the predicted recombination sites at 2100 and 3100 (see Fig. 4). The two genes share the 947 bps 99% identical from position 2045 to 2992 in *msg*79 which are underlined.