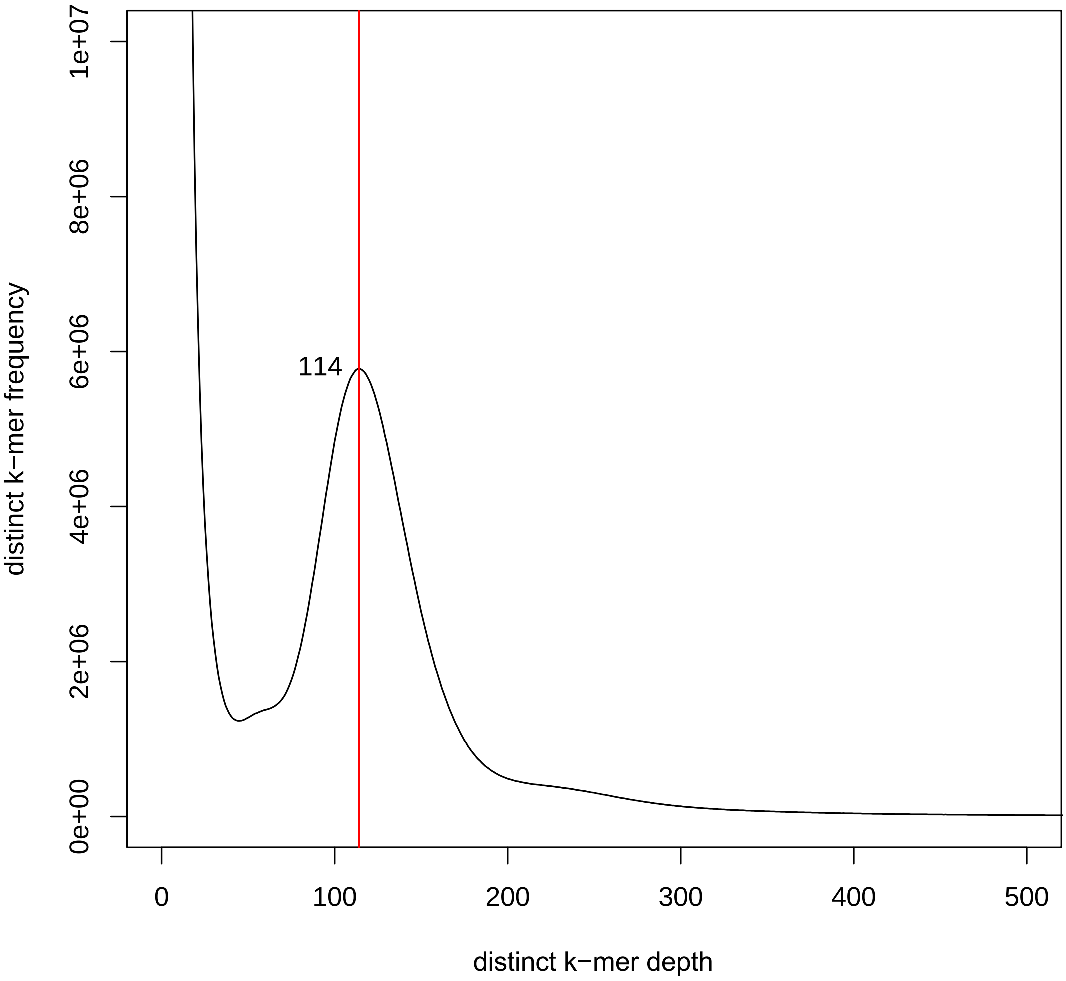
Supplementary Information

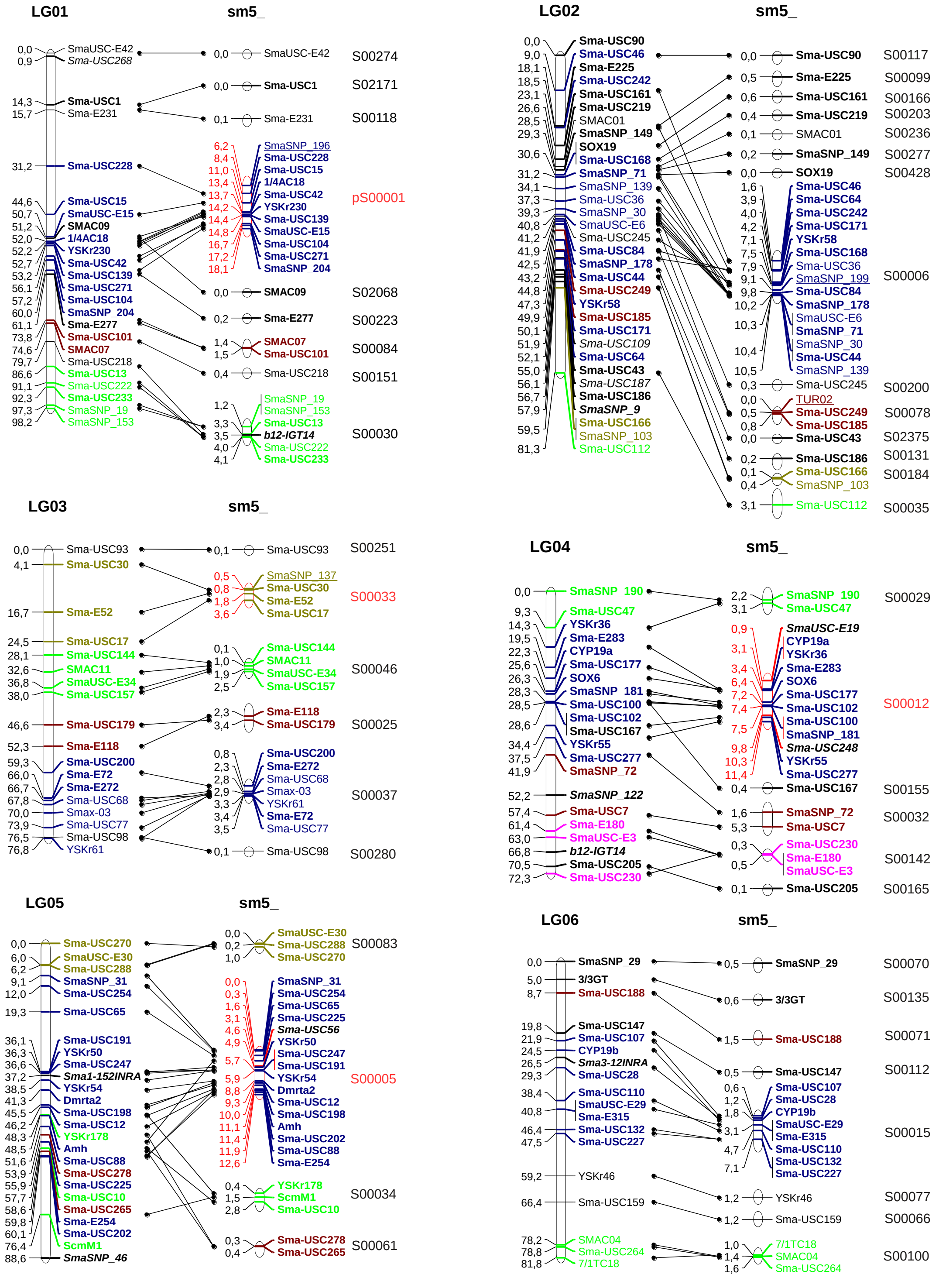
Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life

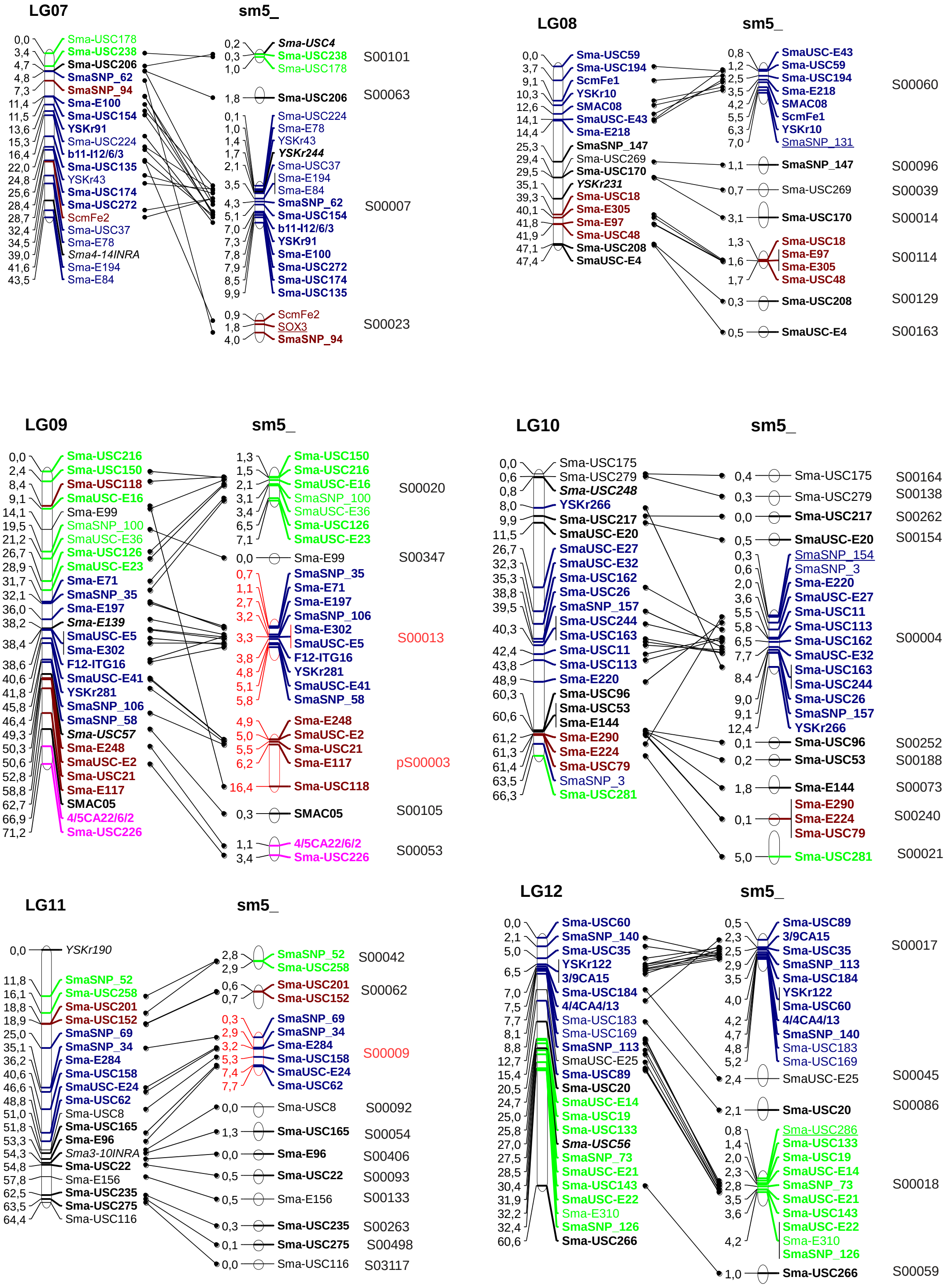
Antonio Figueras[1,*], Diego Robledo[2], André Corvelo[3,4], Miguel Hermida[5], Patricia Pereiro[1], Juan A Rubiolo[5], Jèssica Gómez-Garrido[3], Laia Carreté[3], Xabier Bello[6,7], Marta Gut[3], Ivo Glynne Gut[3], Marina Marcet-Houben[3], Gabriel Forn-Cuní[1], Beatriz Galán[8], José Luis García[8], José Luis Abal-Fabeiro[6,7], Belen G Pardo[5], Xoana Taboada[2], Carlos Fernández[5], Anna Vlasova[3], Antonio Hermoso-Pulido[3], Roderic Guigó[3], José Antonio Álvarez-Dios[9], Antonio Gómez-Tato[10], Ana Viñas[2], Xulio Maside[6,7], Toni Gabaldón[3,11,12], Beatriz Novoa[1], Carmen Bouza[5], Tyler Alioto[3], Paulino Martínez[5,*]

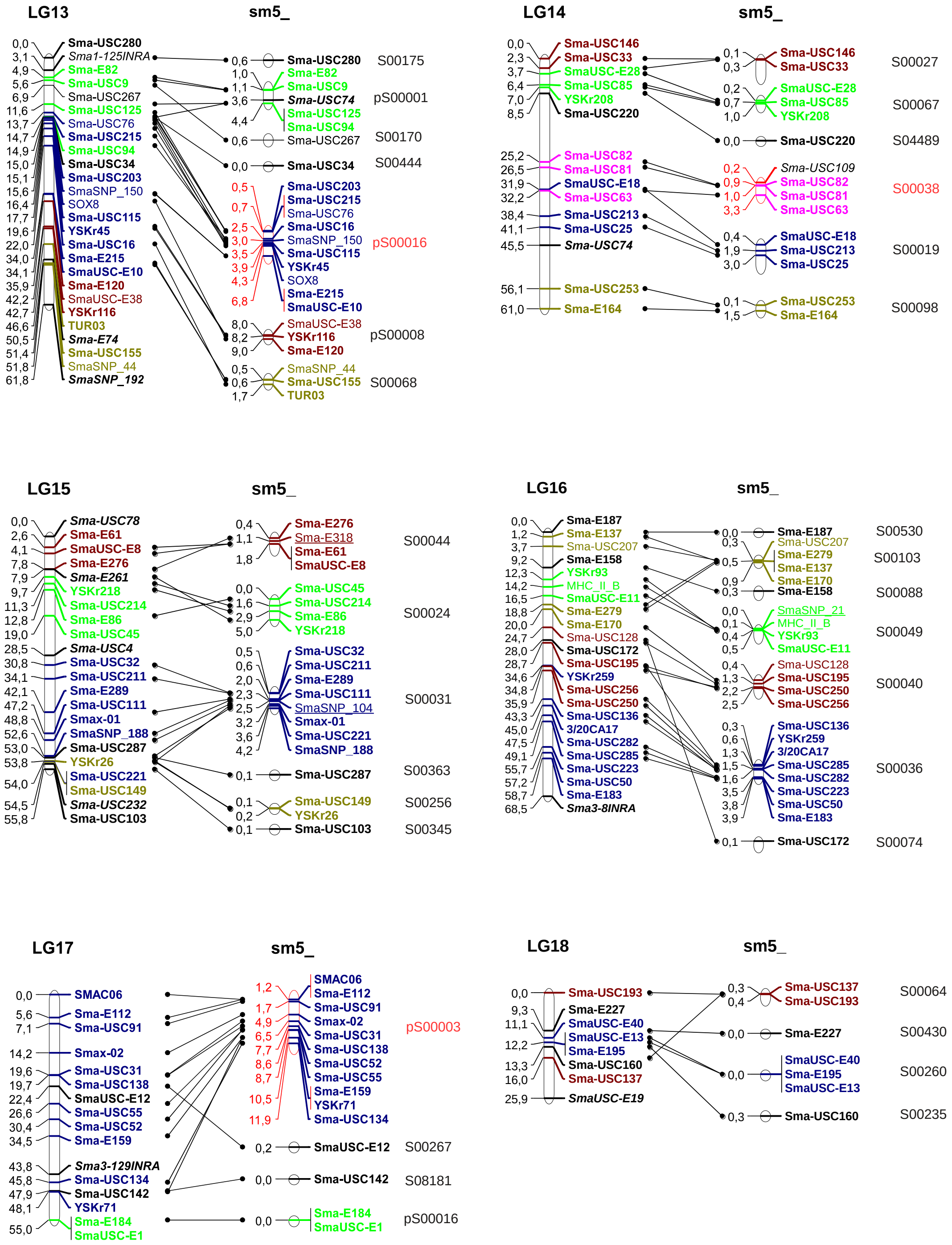
Supplementary Figures

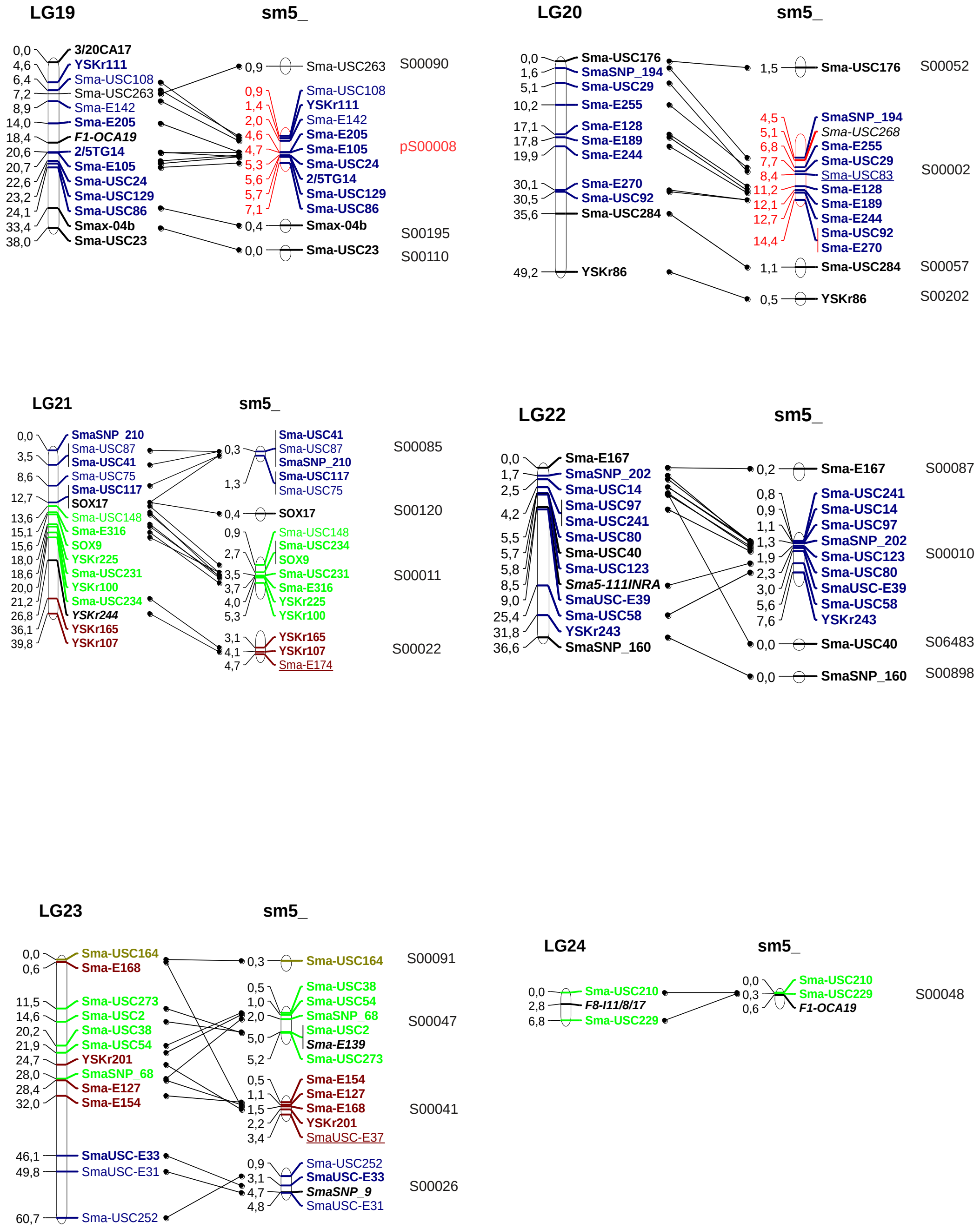


Supplementary Fig. S1. **Estimation of turbot (*S. maximus*) genome size.** Evaluation was performed through a kmer content analysis of the genome using the software Jellyfish v1.1.10[1]. The number of distinct 17-mers are plotted as a function of k-mer multiplicity (depth).

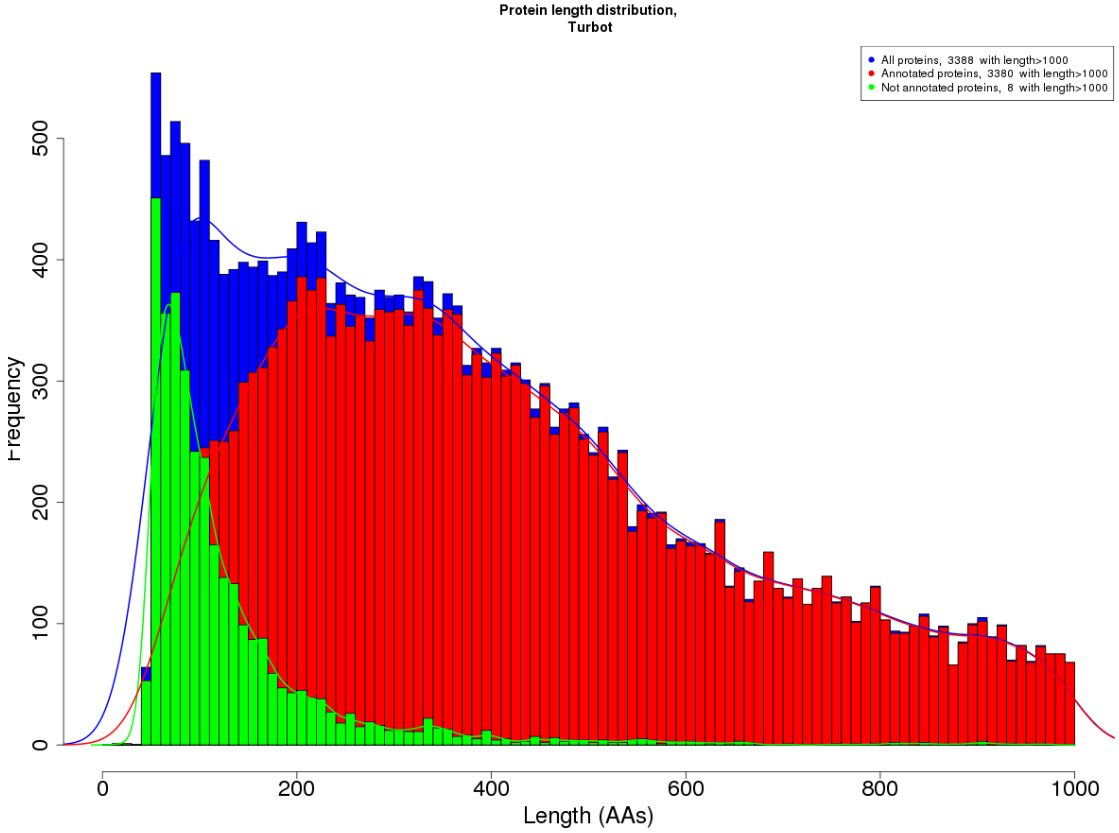

LG01
sm5_
0,0 SmaUSC-E42
0,9 Sma-USC268
14,3 Sma-USC1
15,7 Sma-E231
31,2 Sma-USC228
44,6 Sma-USC15
50,7 SmaUSC-E15
51,2 SMAC09
52,0 1/4AC18
52,2 YSKr230
52,7 Sma-USC42
53,2 Sma-USC139
56,1 Sma-USC271
57,2 Sma-USC104
60,0 SmaSNP_204
61,1 Sma-E277
73,8 Sma-USC101
74,6 SMAC07
79,7 Sma-USC218
86,6 Sma-USC13
91,1 Sma-USC222
92,3 Sma-USC233
97,3 SmaSNP_19
98,2 SmaSNP_153
0,0 SmaUSC-E42 S00274
0,0 Sma-USC1 S02171
0,1 Sma-E231 S00118
6,2 SmaSNP_196
8,4 Sma-USC228
11,0 Sma-USC15
13,4 1/4AC18
13,7 Sma-USC42
14,2 YSKr230
14,4 Sma-USC139
14,8 SmaUSC-E15
16,7 Sma-USC104
17,2 Sma-USC271
18,1 SmaSNP_204
pS00001
0,0 SMAC09 S02068
0,2 Sma-E277 S00223
1,4 SMAC07
1,5 Sma-USC101 S00084
0,4 Sma-USC218 S00151
1,2 SmaSNP_19
SmaSNP_153
3,3 Sma-USC13
3,5 b12-IGT14
4,0 Sma-USC222
4,1 Sma-USC233
S00030
LG02
sm5_
0,0 Sma-USC90
9,0 Sma-USC46
18,1 Sma-E225
18,5 Sma-USC242
23,1 Sma-USC161
26,6 Sma-USC219
28,5 SMAC01
29,3 SmaSNP_149
30,6 SOX19
Sma-USC168
31,2 SmaSNP_71
34,1 SmaSNP_139
37,3 Sma-USC36
39,3 SmaSNP_30
40,8 SmaUSC-E6
41,2 Sma-USC245
41,9 Sma-USC84
42,5 SmaSNP_178
43,2 Sma-USC44
44,8 Sma-USC249
47,3 YSKr58
49,9 Sma-USC185
50,1 Sma-USC171
51,9 Sma-USC109
52,1 Sma-USC64
55,0 Sma-USC43
56,1 Sma-USC187
56,7 Sma-USC186
57,9 SmaSNP_9
59,5 Sma-USC166
SmaSNP_103
81,3 Sma-USC112
0,0 Sma-USC90 S00117
0,5 Sma-E225 S00099
0,6 Sma-USC161 S00166
0,4 Sma-USC219 S00203
0,1 SMAC01 S00236
0,2 SmaSNP_149 S00277
0,0 SOX19 S00428
1,6 Sma-USC46
3,9 Sma-USC64
4,0 Sma-USC242
4,2 Sma-USC171
7,1 YSKr58
7,5 Sma-USC168
7,9 Sma-USC36
9,1 SmaSNP_199
9,8 Sma-USC84
10,2 SmaSNP_178
SmaUSC-E6
10,3 SmaSNP_71
SmaSNP_30
10,4 Sma-USC44
10,5 SmaSNP_139
S00006
0,3 Sma-USC245 S00200
0,0 TUR02
0,5 Sma-USC249
0,8 Sma-USC185
S00078
0,0 Sma-USC43 S02375
0,2 Sma-USC186 S00131
0,1 Sma-USC166
0,4 SmaSNP_103
S00184
3,1 Sma-USC112 S00035
LG03
sm5_
0,0 Sma-USC93
4,1 Sma-USC30
16,7 Sma-E52
24,5 Sma-USC17
28,1 Sma-USC144
32,6 SMAC11
36,8 SmaUSC-E34
38,0 Sma-USC157
46,6 Sma-USC179
52,3 Sma-E118
59,3 Sma-USC200
66,0 Sma-E72
66,7 Sma-E272
67,8 Sma-USC68
70,0 Smax-03
73,9 Sma-USC77
76,5 Sma-USC98
76,8 YSKr61
0,1 Sma-USC93 S00251
0,5 SmaSNP_137
0,8 Sma-USC30
1,8 Sma-E52
3,6 Sma-USC17
S00033
0,1 Sma-USC144
1,0 SMAC11
1,9 SmaUSC-E34
2,5 Sma-USC157
S00046
2,3 Sma-E118
3,4 Sma-USC179
S00025
0,8 Sma-USC200
2,3 Sma-E272
2,8 Sma-USC68
2,9 Smax-03
3,3 YSKr61
3,4 Sma-E72
3,5 Sma-USC77
S00037
0,1 Sma-USC98 S00280
LG04
sm5_
0,0 SmaSNP_190
9,3 Sma-USC47
14,3 YSKr36
19,5 Sma-E283
22,3 CYP19a
25,6 Sma-USC177
26,3 SOX6
28,3 SmaSNP_181
28,5 Sma-USC100
Sma-USC102
28,6 Sma-USC167
34,4 YSKr55
37,5 Sma-USC277
41,9 SmaSNP_72
52,2 SmaSNP_122
57,4 Sma-USC7
61,4 Sma-E180
63,0 SmaUSC-E3
66,8 b12-IGT14
70,5 Sma-USC205
72,3 Sma-USC230
2,2 SmaSNP_190
3,1 Sma-USC47
S00029
0,9 SmaUSC-E19
3,1 CYP19a
YSKr36
3,4 Sma-E283
6,4 SOX6
7,2 Sma-USC177
7,4 Sma-USC102
7,5 Sma-USC100
SmaSNP_181
9,8 Sma-USC248
10,3 YSKr55
11,4 Sma-USC277
S00012
0,4 Sma-USC167 S00155
1,6 SmaSNP_72
5,3 Sma-USC7
S00032
0,3 Sma-USC230
0,5 Sma-E180
SmaUSC-E3
S00142
0,1 Sma-USC205 S00165
LG05
sm5_
0,0 Sma-USC270
6,0 SmaUSC-E30
6,2 Sma-USC288
9,1 SmaSNP_31
12,0 Sma-USC254
19,3 Sma-USC65
36,1 Sma-USC191
36,3 YSKr50
36,6 Sma-USC247
37,2 Sma1-152INRA
38,5 YSKr54
41,3 Dmrta2
45,5 Sma-USC198
46,2 Sma-USC12
48,3 YSKr178
48,5 Amh
51,6 Sma-USC88
53,9 Sma-USC278
55,9 Sma-USC225
57,7 Sma-USC10
58,6 Sma-USC265
59,8 Sma-E254
60,1 Sma-USC202
76,4 ScmM1
88,6 SmaSNP_46
0,0 SmaUSC-E30
0,2 Sma-USC288
1,0 Sma-USC270
S00083
0,0 SmaSNP_31
0,3 Sma-USC254
1,6 Sma-USC65
3,1 Sma-USC225
4,6 Sma-USC56
4,9 YSKr50
5,7 Sma-USC247
Sma-USC191
5,9 YSKr54
8,8 Dmrta2
9,3 Sma-USC12
10,0 Sma-USC198
11,1 Amh
11,4 Sma-USC202
11,9 Sma-USC88
12,6 Sma-E254
S00005
0,4 YSKr178
1,5 ScmM1
2,8 Sma-USC10
S00034
0,3 Sma-USC278
0,4 Sma-USC265
S00061
LG06
sm5_
0,0 SmaSNP_29
5,0 3/3GT
8,7 Sma-USC188
19,8 Sma-USC147
21,9 Sma-USC107
24,5 CYP19b
26,5 Sma3-12INRA
29,3 Sma-USC28
38,4 Sma-USC110
40,8 SmaUSC-E29
Sma-E315
46,4 Sma-USC132
47,5 Sma-USC227
59,2 YSKr46
66,4 Sma-USC159
78,2 SMAC04
78,8 Sma-USC264
81,8 7/1TC18
0,5 SmaSNP_29 S00070
0,6 3/3GT S00135
1,5 Sma-USC188 S00071
0,5 Sma-USC147 S00112
0,6 Sma-USC107
1,2 Sma-USC28
1,8 CYP19b
SmaUSC-E29
3,1 Sma-E315
4,7 Sma-USC110
7,1 Sma-USC132
Sma-USC227
S00015
1,2 YSKr46 S00077
1,2 Sma-USC159 S00066
1,0 7/1TC18
1,4 SMAC04
1,6 Sma-USC264
S00100


LG07
sm5_
0,0 Sma-USC178
3,4 Sma-USC238
4,7 Sma-USC206
4,8 SmaSNP_62
7,3 SmaSNP_94
11,4 Sma-E100
11,5 Sma-USC154
13,6 YSKr91
15,3 Sma-USC224
16,4 b11-I12/6/3
22,0 Sma-USC135
24,8 YSKr43
25,6 Sma-USC174
28,4 Sma-USC272
28,7 ScmFe2
32,4 Sma-USC37
34,5 Sma-E78
39,0 Sma4-14INRA
41,6 Sma-E194
43,5 Sma-E84
0,2 Sma-USC4
0,3 Sma-USC238
1,0 Sma-USC178
S00101
1,8 Sma-USC206
S00063
0,1 Sma-USC224
1,0 Sma-E78
1,4 YSKr43
1,7 YSKr244
2,1 Sma-USC37
Sma-E194
3,5 Sma-E84
4,3 SmaSNP_62
5,1 Sma-USC154
7,0 b11-I12/6/3
7,3 YSKr91
7,8 Sma-E100
7,9 Sma-USC272
8,5 Sma-USC174
9,9 Sma-USC135
S00007
0,9 ScmFe2
1,8 SOX3
4,0 SmaSNP_94
S00023
LG08
sm5_
0,0 Sma-USC59
3,7 Sma-USC194
9,1 ScmFe1
10,3 YSKr10
12,6 SMAC08
14,1 SmaUSC-E43
14,4 Sma-E218
25,3 SmaSNP_147
29,4 Sma-USC269
29,5 Sma-USC170
35,1 YSKr231
39,3 Sma-USC18
40,1 Sma-E305
41,8 Sma-E97
41,9 Sma-USC48
47,1 Sma-USC208
47,4 SmaUSC-E4
0,8 SmaUSC-E43
1,2 Sma-USC59
2,5 Sma-USC194
3,5 Sma-E218
4,2 SMAC08
5,5 ScmFe1
6,3 YSKr10
7,0 SmaSNP_131
S00060
1,1 SmaSNP_147
S00096
0,7 Sma-USC269
S00039
3,1 Sma-USC170
S00014
1,3 Sma-USC18
1,6 Sma-E97
Sma-E305
1,7 Sma-USC48
S00114
0,3 Sma-USC208
S00129
0,5 SmaUSC-E4
S00163
LG09
sm5_
0,0 Sma-USC216
2,4 Sma-USC150
8,4 Sma-USC118
9,1 SmaUSC-E16
14,1 Sma-E99
19,5 SmaSNP_100
21,2 SmaUSC-E36
26,7 Sma-USC126
28,9 SmaUSC-E23
31,7 Sma-E71
32,1 SmaSNP_35
36,0 Sma-E197
38,2 Sma-E139
38,4 SmaUSC-E5
Sma-E302
38,6 F12-ITG16
40,6 SmaUSC-E41
41,8 YSKr281
45,8 SmaSNP_106
46,4 SmaSNP_58
49,3 Sma-USC57
50,3 Sma-E248
50,6 SmaUSC-E2
52,8 Sma-USC21
58,8 Sma-E117
62,7 SMAC05
66,9 4/5CA22/6/2
71,2 Sma-USC226
1,3 Sma-USC150
1,5 Sma-USC216
2,1 SmaUSC-E16
3,1 SmaSNP_100
3,4 SmaUSC-E36
6,5 Sma-USC126
7,1 SmaUSC-E23
S00020
0,0 Sma-E99
S00347
0,7 SmaSNP_35
1,1 Sma-E71
2,7 Sma-E197
3,2 SmaSNP_106
3,3 Sma-E302
SmaUSC-E5
3,8 F12-ITG16
4,8 YSKr281
5,1 SmaUSC-E41
5,8 SmaSNP_58
S00013
4,9 Sma-E248
5,0 SmaUSC-E2
5,5 Sma-USC21
6,2 Sma-E117
pS00003
16,4 Sma-USC118
0,3 SMAC05
S00105
1,1 4/5CA22/6/2
3,4 Sma-USC226
S00053
LG10
sm5_
0,0 Sma-USC175
0,6 Sma-USC279
0,8 Sma-USC248
8,0 YSKr266
9,9 Sma-USC217
11,5 SmaUSC-E20
26,7 SmaUSC-E27
32,3 SmaUSC-E32
35,3 Sma-USC162
38,8 Sma-USC26
39,5 SmaSNP_157
40,3 Sma-USC244
Sma-USC163
42,4 Sma-USC11
43,8 Sma-USC113
48,9 Sma-E220
60,3 Sma-USC96
60,6 Sma-USC53
Sma-E144
61,2 Sma-E290
61,3 Sma-E224
61,4 Sma-USC79
63,5 SmaSNP_3
66,3 Sma-USC281
0,4 Sma-USC175
S00164
0,3 Sma-USC279
S00138
0,0 Sma-USC217
S00262
0,5 SmaUSC-E20
S00154
0,3 SmaSNP_154
0,6 SmaSNP_3
2,0 Sma-E220
3,6 SmaUSC-E27
5,5 Sma-USC11
5,8 Sma-USC113
6,5 Sma-USC162
7,7 SmaUSC-E32
8,4 Sma-USC163
Sma-USC244
9,0 Sma-USC26
9,1 SmaSNP_157
12,4 YSKr266
S00004
0,1 Sma-USC96
S00252
0,2 Sma-USC53
S00188
1,8 Sma-E144
S00073
Sma-E290
0,1 Sma-E224
Sma-USC79
S00240
5,0 Sma-USC281
S00021
LG11
sm5_
0,0 YSKr190
11,8 SmaSNP_52
16,1 Sma-USC258
18,8 Sma-USC201
18,9 Sma-USC152
25,0 SmaSNP_69
35,1 SmaSNP_34
36,2 Sma-E284
40,6 Sma-USC158
46,6 SmaUSC-E24
48,8 Sma-USC62
51,0 Sma-USC8
51,8 Sma-USC165
53,3 Sma-E96
54,3 Sma3-10INRA
54,8 Sma-USC22
57,8 Sma-E156
62,5 Sma-USC235
63,5 Sma-USC275
64,4 Sma-USC116
2,8 SmaSNP_52
2,9 Sma-USC258
S00042
0,6 Sma-USC201
0,7 Sma-USC152
S00062
0,3 SmaSNP_69
2,9 SmaSNP_34
3,2 Sma-E284
5,3 Sma-USC158
7,4 SmaUSC-E24
7,7 Sma-USC62
S00009
0,0 Sma-USC8
S00092
1,3 Sma-USC165
S00054
0,0 Sma-E96
S00406
0,5 Sma-USC22
S00093
0,5 Sma-E156
S00133
0,3 Sma-USC235
S00263
0,1 Sma-USC275
S00498
0,0 Sma-USC116
S03117
LG12
sm5_
0,0 Sma-USC60
2,1 SmaSNP_140
5,0 Sma-USC35
6,5 YSKr122
3/9CA15
7,0 Sma-USC184
7,5 4/4CA4/13
7,7 Sma-USC183
8,1 Sma-USC169
8,8 SmaSNP_113
12,7 SmaUSC-E25
15,4 Sma-USC89
20,5 Sma-USC20
24,7 SmaUSC-E14
25,0 Sma-USC19
25,8 Sma-USC133
27,0 Sma-USC56
27,5 SmaSNP_73
28,5 SmaUSC-E21
30,4 Sma-USC143
31,9 SmaUSC-E22
32,2 Sma-E310
32,4 SmaSNP_126
60,6 Sma-USC266
0,5 Sma-USC89
2,3 3/9CA15
2,5 Sma-USC35
2,9 SmaSNP_113
3,5 Sma-USC184
4,0 YSKr122
Sma-USC60
4,2 4/4CA4/13
4,7 SmaSNP_140
4,8 Sma-USC183
5,2 Sma-USC169
S00017
2,4 SmaUSC-E25
S00045
2,1 Sma-USC20
S00086
0,8 Sma-USC286
1,4 Sma-USC133
2,0 Sma-USC19
2,3 SmaUSC-E14
2,8 SmaSNP_73
3,5 SmaUSC-E21
3,6 Sma-USC143
SmaUSC-E22
4,2 Sma-E310
SmaSNP_126
S00018
1,0 Sma-USC266
S00059

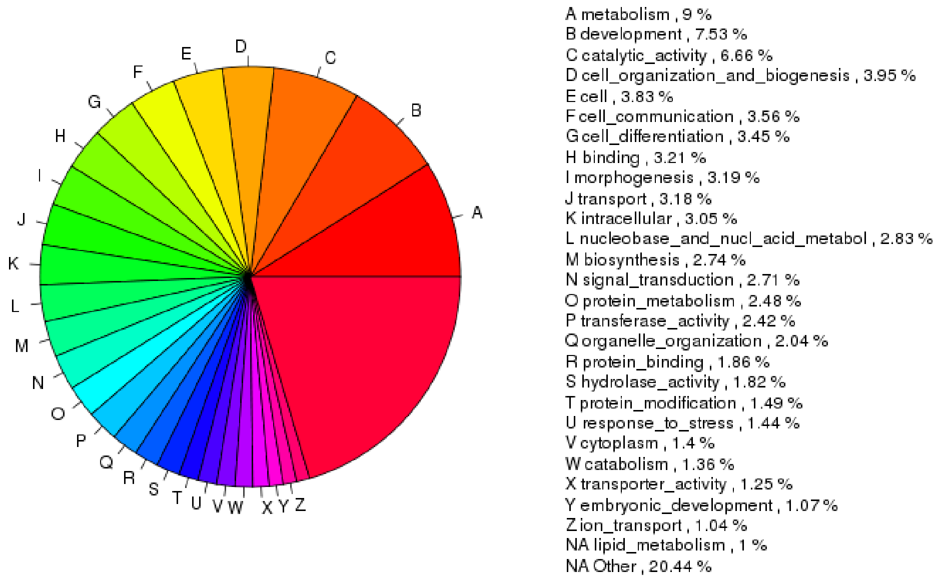

LG13
sm5_
0,0 Sma-USC280
3,1 Sma1-125INRA
4,9 Sma-E82
5,6 Sma-USC9
6,9 Sma-USC267
11,6 Sma-USC125
13,7 Sma-USC76
14,7 Sma-USC215
14,9 Sma-USC94
15,0 Sma-USC34
15,1 Sma-USC203
15,6 SmaSNP_150
16,4 SOX8
17,7 Sma-USC115
19,6 YSKr45
22,0 Sma-USC16
34,0 Sma-E215
34,1 SmaUSC-E10
35,9 Sma-E120
42,2 SmaUSC-E38
42,7 YSKr116
46,6 TUR03
50,5 Sma-E74
51,4 Sma-USC155
51,8 SmaSNP_44
61,8 SmaSNP_192
0,6 Sma-USC280
S00175
1,0 Sma-E82
1,1 Sma-USC9
3,6 Sma-USC74
4,4 Sma-USC125
Sma-USC94
pS00001
0,6 Sma-USC267
S00170
0,0 Sma-USC34
S00444
0,5 Sma-USC203
0,7 Sma-USC215
Sma-USC76
2,5 Sma-USC16
3,0 SmaSNP_150
3,5 Sma-USC115
3,9 YSKr45
4,3 SOX8
6,8 Sma-E215
SmaUSC-E10
pS00016
8,0 SmaUSC-E38
8,2 YSKr116
9,0 Sma-E120
pS00008
0,5 SmaSNP_44
0,6 Sma-USC155
1,7 TUR03
S00068
LG14
sm5_
0,0 Sma-USC146
2,3 Sma-USC33
3,7 SmaUSC-E28
6,4 Sma-USC85
7,0 YSKr208
8,5 Sma-USC220
25,2 Sma-USC82
26,5 Sma-USC81
31,9 SmaUSC-E18
32,2 Sma-USC63
38,4 Sma-USC213
41,1 Sma-USC25
45,5 Sma-USC74
56,1 Sma-USC253
61,0 Sma-E164
0,1 Sma-USC146
0,3 Sma-USC33
S00027
0,2 SmaUSC-E28
0,7 Sma-USC85
1,0 YSKr208
S00067
0,0 Sma-USC220
S04489
0,2 Sma-USC109
0,9 Sma-USC82
1,0 Sma-USC81
3,3 Sma-USC63
S00038
0,4 SmaUSC-E18
1,9 Sma-USC213
3,0 Sma-USC25
S00019
0,1 Sma-USC253
1,5 Sma-E164
S00098
LG15
sm5_
0,0 Sma-USC78
2,6 Sma-E61
4,1 SmaUSC-E8
7,8 Sma-E276
7,9 Sma-E261
9,7 YSKr218
11,3 Sma-USC214
12,8 Sma-E86
19,0 Sma-USC45
28,5 Sma-USC4
30,8 Sma-USC32
34,1 Sma-USC211
42,1 Sma-E289
47,2 Sma-USC111
48,8 Smax-01
52,6 SmaSNP_188
53,0 Sma-USC287
53,8 YSKr26
54,0 Sma-USC221
Sma-USC149
54,5 Sma-USC232
55,8 Sma-USC103
0,4 Sma-E276
1,1 Sma-E318
1,8 Sma-E61
SmaUSC-E8
S00044
0,0 Sma-USC45
1,6 Sma-USC214
2,9 Sma-E86
5,0 YSKr218
S00024
0,5 Sma-USC32
0,6 Sma-USC211
2,0 Sma-E289
2,3 Sma-USC111
2,5 SmaSNP_104
3,2 Smax-01
3,6 Sma-USC221
4,2 SmaSNP_188
S00031
0,1 Sma-USC287
S00363
0,1 Sma-USC149
0,2 YSKr26
S00256
0,1 Sma-USC103
S00345
LG16
sm5_
0,0 Sma-E187
1,2 Sma-E137
3,7 Sma-USC207
9,2 Sma-E158
12,3 YSKr93
14,2 MHC_II_B
16,5 SmaUSC-E11
18,8 Sma-E279
20,0 Sma-E170
24,7 Sma-USC128
28,0 Sma-USC172
28,7 Sma-USC195
34,6 YSKr259
34,8 Sma-USC256
35,9 Sma-USC250
43,3 Sma-USC136
45,0 3/20CA17
47,5 Sma-USC282
49,1 Sma-USC285
55,7 Sma-USC223
57,2 Sma-USC50
58,7 Sma-E183
68,5 Sma3-8INRA
0,0 Sma-E187
S00530
0,3 Sma-USC207
0,5 Sma-E279
Sma-E137
0,9 Sma-E170
S00103
0,3 Sma-E158
S00088
0,0 SmaSNP_21
0,1 MHC_II_B
0,4 YSKr93
0,5 SmaUSC-E11
S00049
0,4 Sma-USC128
1,3 Sma-USC195
2,2 Sma-USC250
2,5 Sma-USC256
S00040
0,3 Sma-USC136
0,6 YSKr259
1,3 3/20CA17
1,5 Sma-USC285
1,6 Sma-USC282
3,5 Sma-USC223
3,8 Sma-USC50
3,9 Sma-E183
S00036
0,1 Sma-USC172
S00074
LG17
sm5_
0,0 SMAC06
5,6 Sma-E112
7,1 Sma-USC91
14,2 Smax-02
19,6 Sma-USC31
19,7 Sma-USC138
22,4 SmaUSC-E12
26,6 Sma-USC55
30,4 Sma-USC52
34,5 Sma-E159
43,8 Sma3-129INRA
45,8 Sma-USC134
47,9 Sma-USC142
48,1 YSKr71
55,0 Sma-E184
SmaUSC-E1
1,2 SMAC06
1,7 Sma-E112
4,9 Sma-USC91
6,5 Smax-02
7,7 Sma-USC31
8,6 Sma-USC138
8,7 Sma-USC52
10,5 Sma-USC55
11,9 Sma-E159
YSKr71
Sma-USC134
pS00003
0,2 SmaUSC-E12
S00267
0,0 Sma-USC142
S08181
0,0 Sma-E184
SmaUSC-E1
pS00016
LG18
sm5_
0,0 Sma-USC193
9,3 Sma-E227
11,1 SmaUSC-E40
SmaUSC-E13
12,2 Sma-E195
13,3 Sma-USC160
16,0 Sma-USC137
25,9 SmaUSC-E19
0,3 Sma-USC137
0,4 Sma-USC193
S00064
0,0 Sma-E227
S00430
0,0 SmaUSC-E40
Sma-E195
SmaUSC-E13
S00260
0,3 Sma-USC160
S00235



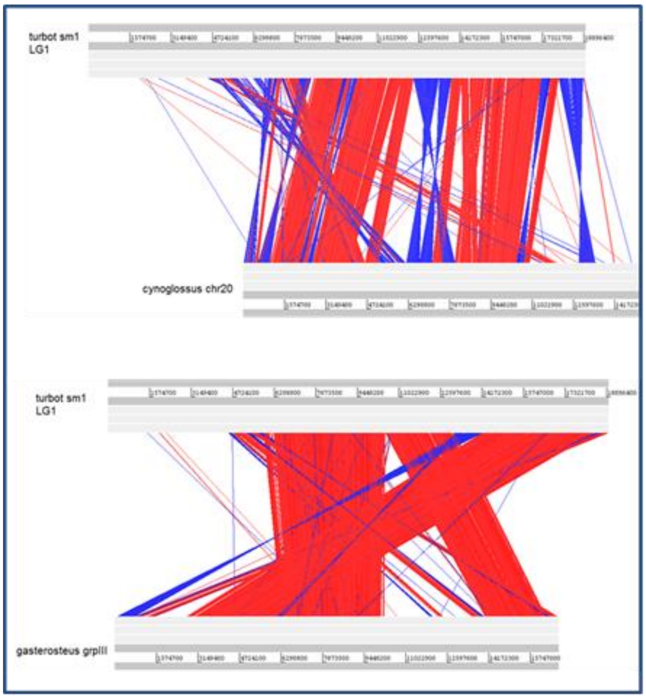
Supplementary Fig. S2. Anchoring of the turbot (*S. maximus*) genome assembly (right) on the turbot genetic map (left). The position of markers in the genetic map and in the scaffolds where their sequences matched is shown. Marker positions are represented in cM in the genetic map and in Mb in the genome. Scaffolds with two or more markers from the same LG are colored (colors are meaningful), whereas scaffolds with only one marker are in black. Framework markers are represented in bold type. Markers of the genetic map with no match in scaffolds, and *viceversa*, are in italic type. The marker closest to the centromere is underlined. The scaffolds with positions in red have been inverted only for representation purposes.

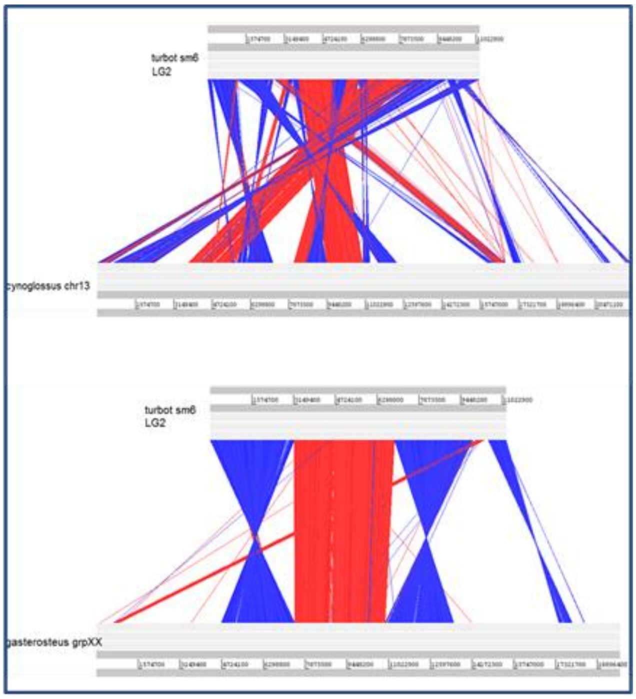


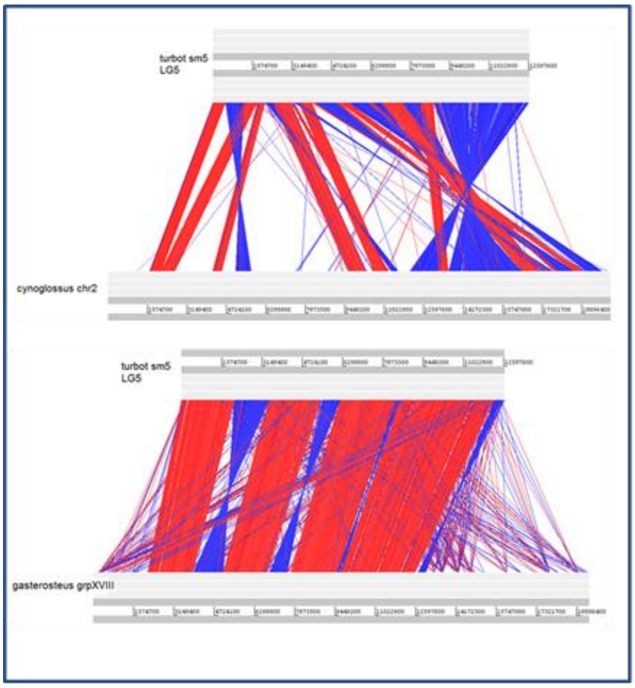
Supplementary Fig. S3. Size distribution and annotation of the turbot (*S. maximus*) transcriptome (protein coding genes). Blue: all mRNAs; red: annotated mRNAs; green: non-annotated mRNAs.

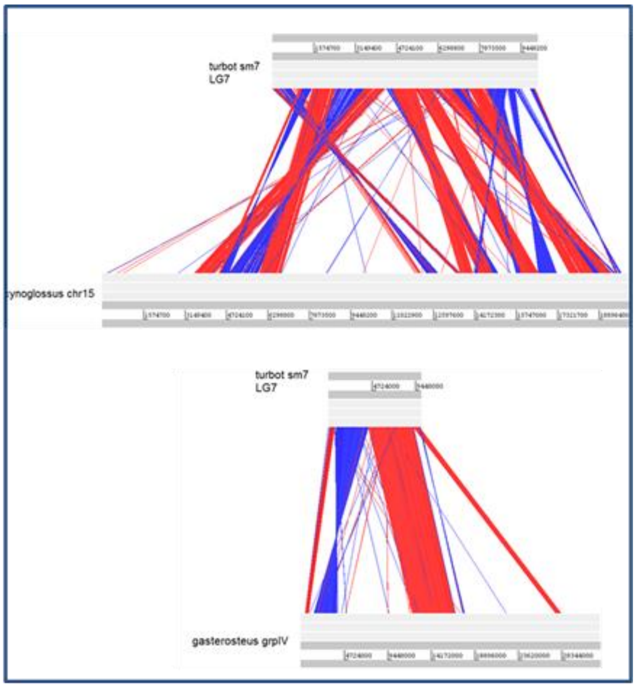


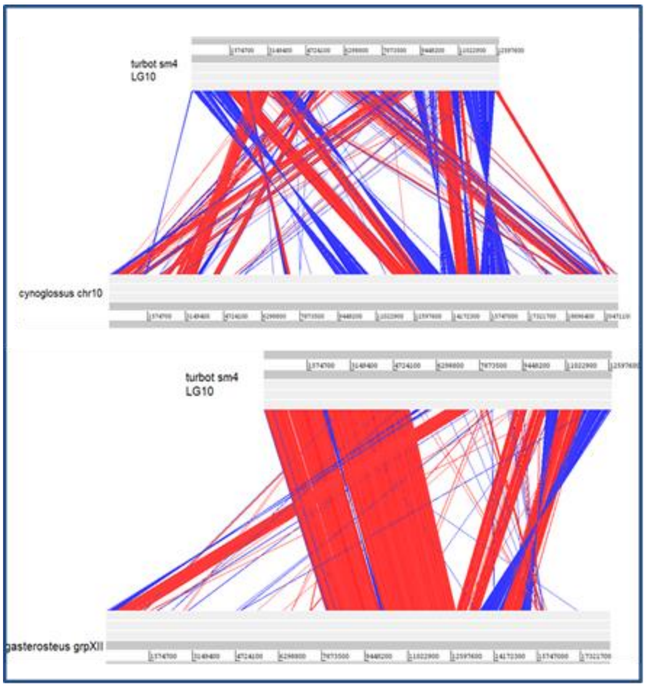
Supplementary Fig. S4. Distribution of gene ontology (GO) functional terms of the turbot (*S. maximus*) proteome.

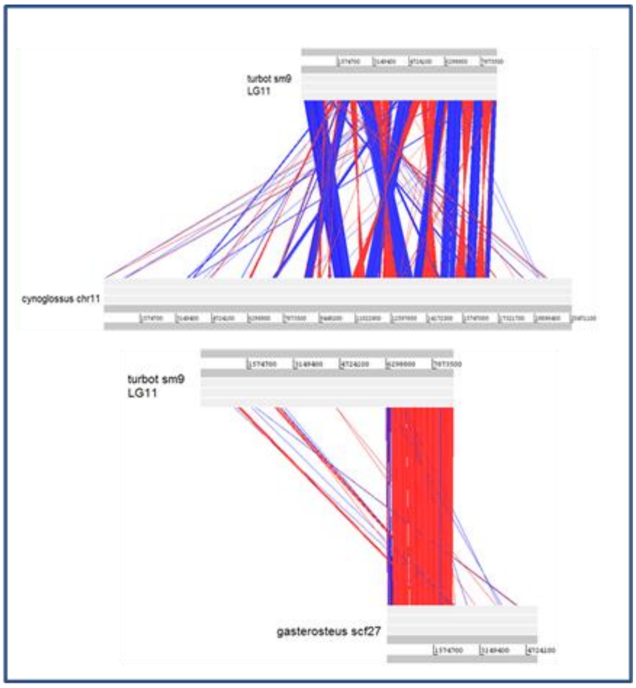

turbot sm1
LG1
cynoglossus chr20
turbot sm1
LG1
gasterosteus grpIII

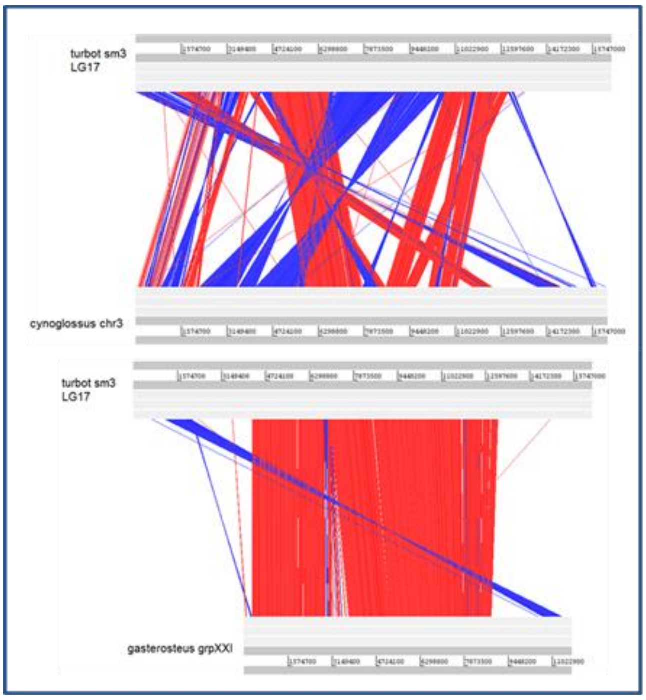

turbot sm6
LG2
cynoglossus chr13
turbot sm6
LG2
gasterosteus grpXX

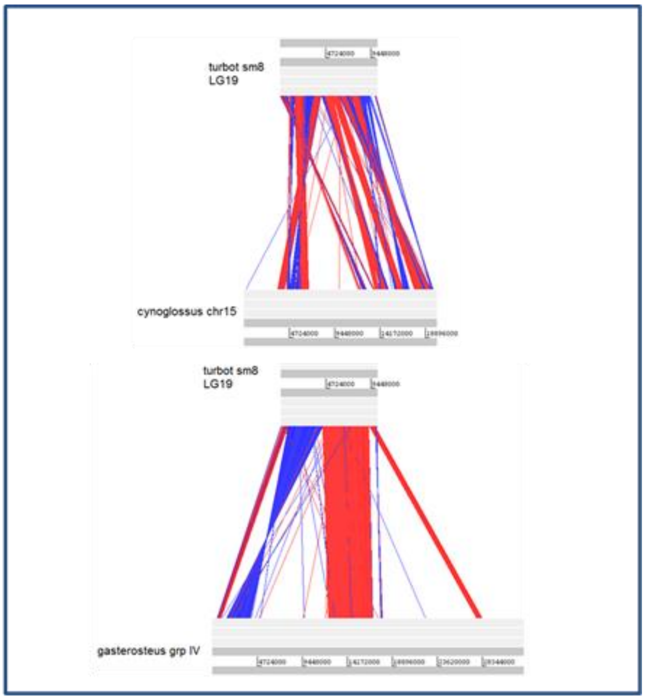

turbot sm5
LG5
cynoglossus chr2
turbot sm5
LG5
gasterosteus grpXVIII

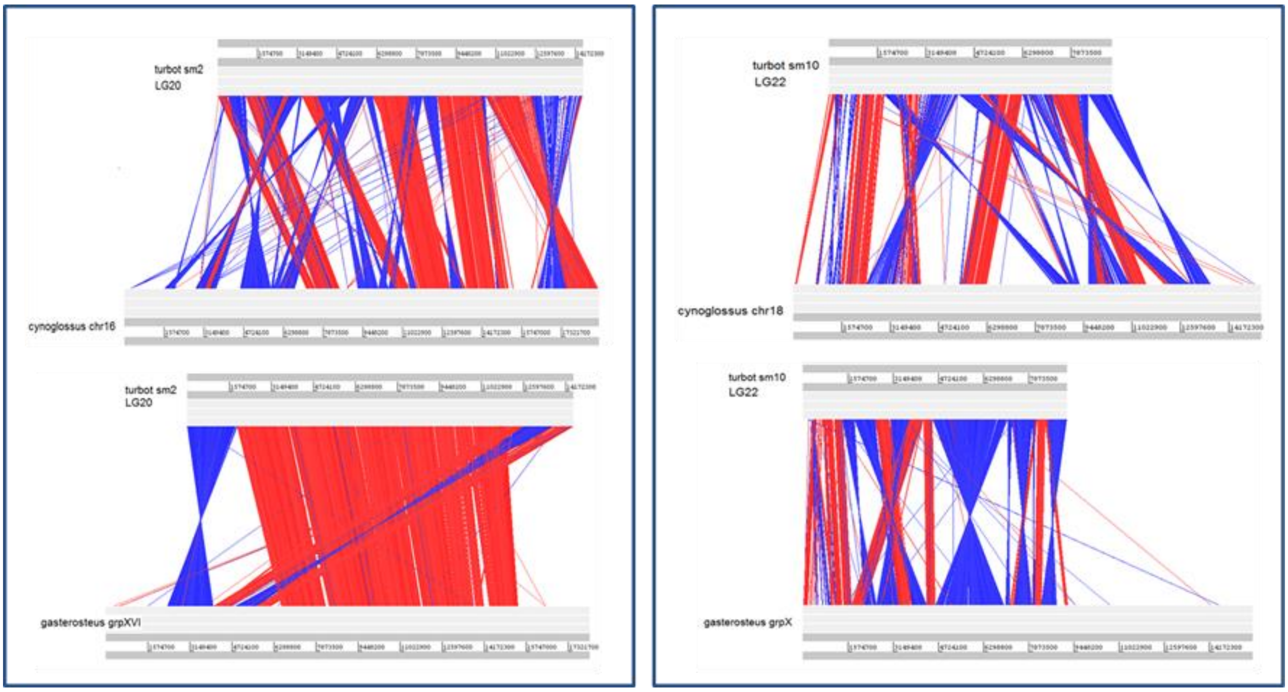

turbot sm7
LG7
cynoglossus chr15
turbot sm7
LG7
gasterosteus grpIV


turbot sm4
LG10
1574700
3149400
4724100
6298800
7873500
9448200
11022900
12597600
cynoglossus chr10
1574700
3149400
4724100
6298800
7873500
9448200
11022900
12597600
14172300
15747000
17321700
18896400
20471100
turbot sm4
LG10
1574700
3149400
4724100
6298800
7873500
9448200
11022900
12597600
gasterosteus grpXII
1574700
3149400
4724100
6298800
7873500
9448200
11022900
12597600
14172300
15747000
17321700

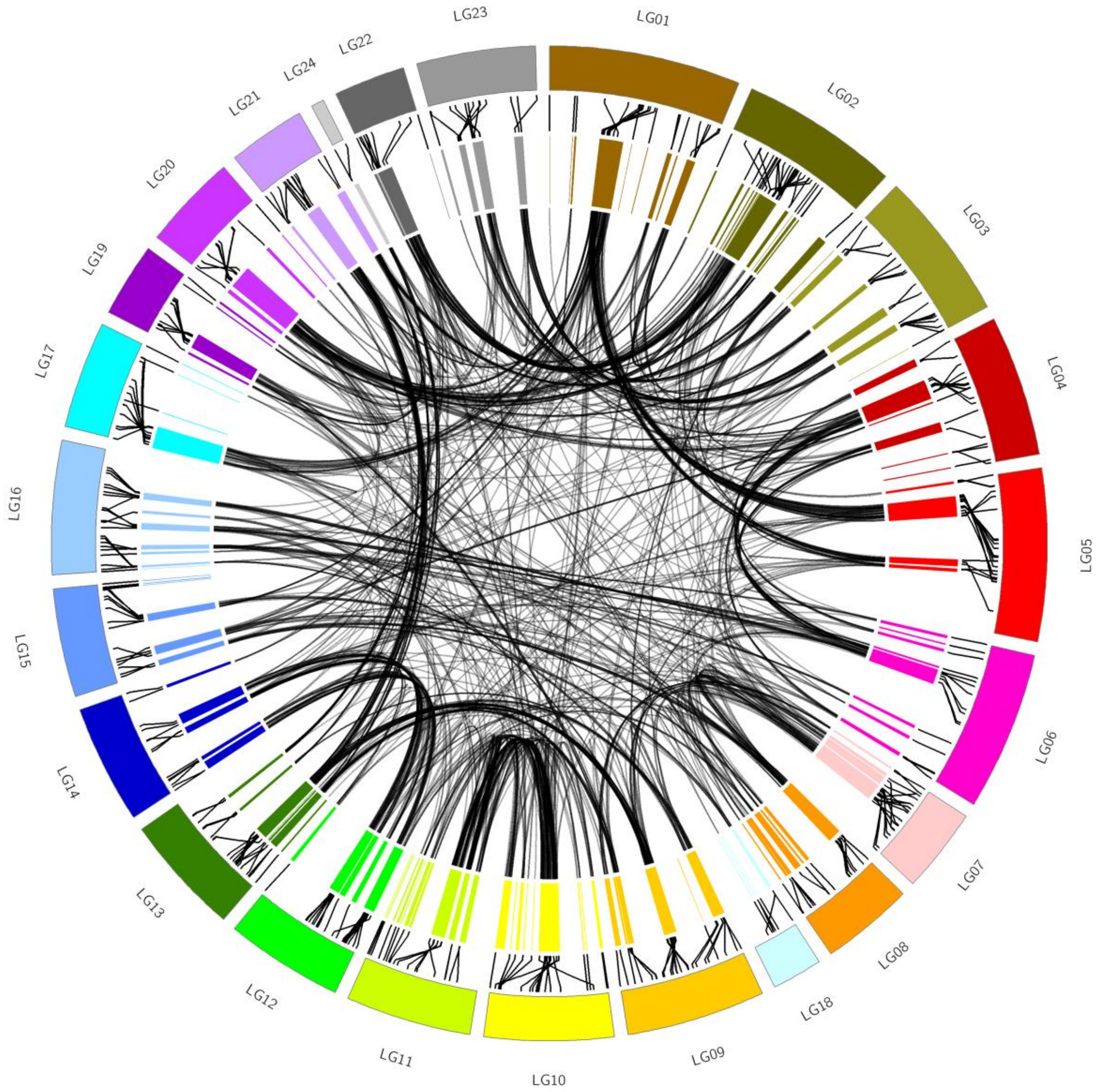

turbot sm9
LG11
1574700
3149400
4724100
6298800
7873500
cynoglossus chr11
1574700
3149400
4724100
6298800
7873500
9448200
11022900
12597600
14172300
15747000
17321700
18896400
20471100
turbot sm9
LG11
1574700
3149400
4724100
6298800
7873500
gasterosteus scf27
1574700
3149400
4724100


turbot sm3
LG17
1574700
3149400
4724100
6298800
7873500
9448200
11022900
12597600
14172300
15747000
cynoglossus chr3
1574700
3149400
4724100
6298800
7873500
9448200
11022900
12597600
14172300
15747000
turbot sm3
LG17
1574700
3149400
4724100
6298800
7873500
9448200
11022900
12597600
14172300
15747000
gasterosteus grpXXI
1574700
3149400
4724100
6298800
7873500
9448200
11022900


turbot sm8
LG19
4724000
9448000
cynoglossus chr15
4724000
9448000
14172000
18896000
turbot sm8
LG19
4724000
9448000
gasterosteus grp IV
4724000
9448000
14172000
18896000
23620000
28344000



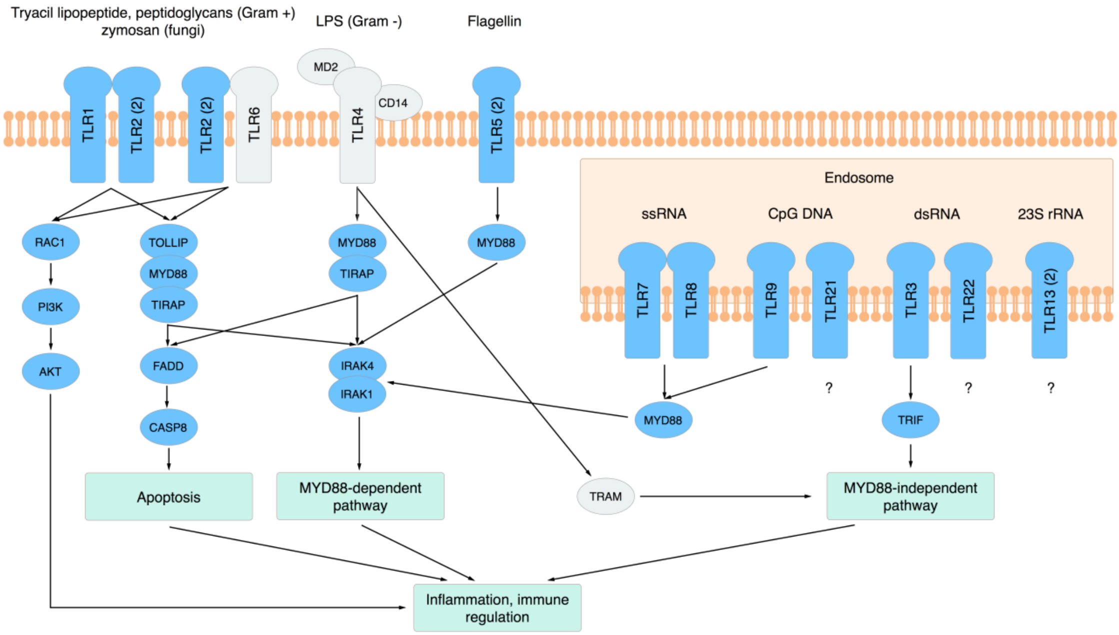
Supplementary Fig. S5. Microreorganizations in the turbot (*S. maximus*) genome when compared with the genomes of the two closest fish species, stickleback (*G. aculeatus*) and tongue sole (*C. semilaevis*). The sequences of the 10 biggest scaffolds from turbot genome were compared (E-100) against the genomes of the closest species. Positions within scaffolds or chromosomes are indicated in bp.



Supplementary Fig. S6. Circle diagram showing the syntenic pattern within the turbot (*S. maximus*) genome from paralogous relationships.

Turbot LG	LG01	LG02	LG03	LG04	LG05	LG06	LG07	LG08+18	LG09	LG10	LG11	LG12	LG13	LG14	LG15	LG16	LG17	LG19	LG20	LG21+24	LG22	LG23
LG01	■																					
LG02	12	■																				
LG03	9	1	■																			
LG04	3	6		■																		
LG05	73	8	2	3	■																	
LG06	1	3		48		■																
LG07	1	2	3	4	1		■															
LG08+18	1	1		3	3	5	25	■														
LG09	1	6	1		4	3	22	1	■													
LG10	8	1		1	1	2			1	■												
LG11	2	1				2	1	1		93	■											
LG12	3	1	3	3	3	1	2	9	3		3	■										
LG13	2		1						78		1	3	■									
LG14	5	1		2	7	1	1	1	2	1		73		■								
LG15	12		3	6	1	2	3	24				1		1	■							
LG16	2	1		1		35		3	10	4	8		1		1	■						
LG17	21		2	1			1	6	2	2			1			1	■					
LG19	1	2	3			1		6	11		1	2		1		1		■				
LG20	12	38	7	12			1	1	2	3	2		1		2	1	3	3	■			
LG21+24	7	19	3				2	5	1	2			60		1	1		15	1	■		
LG22		47		2	1	1		2		3			2	6				1			■	
LG23	1	5	52	8		2		2		3			16	3			1	5	1		4	■

Supplementary Fig. S7. Oxford plot showing paralogous relationships in the turbot (*S. maximus*) genome. In yellow the highest figures for each linkage group (LG).



Supplementary Fig. S8. Toll-like receptor pathways inferred from genome information of turbot (*S. maximus*). Although most of the mammalian TLRs were detected in the turbot genome, three of them (*tlr4*, *tlr6* and *tlr10*) are apparently missing. In agreement with published teleost genomes to date (with the exception of Cyprinidae), tlr4 seems not to exist in teleost. The genes encoding its accessory molecules *cd14* and *md2*, constituting the lipopolysaccharide receptor complex tlr4/cd14/md2, are also missing. These observations suggest an alternative LPS-recognition pathway in fish.

Supplementary Tables

Supplementary Table S1. **List of species used in the phylome reconstruction**. TaxId: taxon identification code.

TaxID	Species name	Source of protein coding sequences
7868	*Callorhinchus milii*	Ensembl_75
7897	*Latimeria chalumnae*	Ensembl_69
7918	*Lepisosteus oculatus*	Ensembl_75
7955	*Danio rerio*	Quest for Orthologs 2012_05
7994	*Astyanax mexicanus*	Ensembl_75
8022	*Oncorhynchus mykiss*	Genoscope
8049	*Gadus morhua*	Ensembl_75
8083	*Xiphophorus maculatus*	Ensembl_75
8090	*Oryzias latipes*	Ensembl_75
8128	*Oreochromis niloticus*	Ensembl_75
8364	*Xenopus tropicalis*	Quest for Orthologs 2012_05
10090	*Mus musculus*	Quest for Orthologs 2012_05
28377	*Anolis carolinensis*	Ensembl_69
31033	*Takifugu rubripes*	Ensembl_75
52904	*Scophthalmus maximus*	Sequencing project
69293	*Gasterosteus aculeatus*	Ensembl_75
99883	*Tetraodon nigroviridis*	Ensembl_75
244447	*Cynoglossus semilaevis*	NCBI

Supplementary Table S2. Characteristics of the libraries constructed for assembling the turbot (*S. maximus*) genome.

Type	Insert size	Read length	Number of clones	Number of reads	Number of bases	Sequencing depth
PE	200 bp	150		142x10^6	42.5 Gb	60x
PE	500 bp	150		142*10^6	42.5 Gb	60x
MP	3.0 kb	100		145*10^6	29.0 Gb	41x
MP	5.0 kb	100		164*10^6	33.0 Gb	47x
FE	40 kb	100	140,000	60*10^6	12.0 Gb	17x
Total				653*10^6	159 Gb	219x

Supplementary Table S3. Summarized (a) and detailed (b) description of TE-derived sequence and other simple repeats in the turbot (*S. maximus*) genome. N.A.- Not available. Superfamilies contributing < 1kb of genomic sequence were not included.

a

Class	Order	Superfamily	% genome
RTs			2.29
	DIRS		0.05
	LINE		1.21
		L2	0.58
		L1	0.20
		RTE	0.17
	LTR		0.62
		Gypsy	0.34
		ERV	0.21
	PLE		0.03
	SINE		0.37
DNA			2.56
	Helitron	Helitron	0.10
	TIR		2.46
		hAT	0.47
		Tc1-Mariner	0.36
		MITEs	1.33
Unclassified			0.15
Total TE-derived			5.0
Othermotifs			
		SmallRNAs	0.03
		Satellites	0.13
		Simple repeats	3.01
		Lowcomplexity	0.35

b

Class	**Order**	***Clade/Superfamily***	**No. fragments**	**Total length (kbp)**	**Genome fraction (%)**
RTs			76621	12136	2.29
	DIRS		2231	274	0.05
		Ngaro	1344	154	0.03
		DIRS	887	120	0.02
	LINE		28674	6436	1.21
		L2	14588	3055	0.58
		L1	5290	1051	0.20
		RTE	3801	908	0.17
	LTR		28206	3291	0.62
		Gypsy	9027	1781	0.34
		ERV1	8540	690	0.13
		ERVK	6018	389	0.07
		Pao	1283	256	0.05
		Copia	96	20	0.00
		ERVL	332	19	0.00
		ERV4	200	12	0.00
		Ginger	32	5	0.00
		ERVL-MaLR	32	2	0.00
	PLE	Penelope	1613	183	0.03
	SINE		15897	1952	0.01
		tRNA-V	6438	1175	0.22
		MIR	3990	458	0.09
		tRNA-Core	3384	441	0.08
		Mermaid	642	55	0.01
		L2	875	50	0.01
		tRNA	606	45	0.01
		tRNA-Core-L2	606	45	0.01
		tRNA-V-CR1	373	37	0.01
		5S-Deu-L2	358	23	0.00
		5S-Sauria-RTE	135	15	0.00
		tRNA-V-Core-L2	122	13	0.00
		tRNA-L2	148	8	0.00
		tRNA-C	126	7	0.00
		B4	63	4	0.00
		ID	43	2	0.00
		tRNA-Deu-L2	35	2	0.00
		7SL	7	1	0.00
		tRNA-RTE	22	1	0.00
DNA			N.A.	13557	2.56

Helitron	Helitron	5573	513	0.10
TIR		60743	13041	2.12
	hAT	31015	2473	0.47
	Tc1-Mariner	10861	1931	0.36
	EnSpm	7360	605	0.11
	Maverick	4183	371	0.07
	PIF-Harbinger	1875	208	0.04
	Kolobok-T2	2213	156	0.03
	Dada	2160	156	0.03
	PiggyBac	825	68	0.01
	Academ	113	17	0.00
	Harbinger	80	8	0.00
	PIF-ISL2EU	23	2	0.00
	MITEs	N.A.	7043	1.33
Unclassified		9102	793	0.15
Total interspersedrepeats		N.A.	26486	5.00
Other motifs	SmallRNAs	2065	156	0.03
	Satellites	5432	664	0.13
	Simple repeats	358944	15954	3.01
	Lowcomplexity	33492	1860	0.35

Supplementary Table S7. Functional annotation sources for the turbot (*S. maximus*) protein-coding gene annotation.

InterPro member database	Number of proteins
PANTHER	22,132
Pfam	21,247
SUPERFAMILY	17,477
Gene3D	16,659
ProSiteProfiles	12,010
SMART	11,279
ProSitePatterns	7,260
PRINTS	6,466
Coils	5,892
TIGRFAM	1,570
PIRSF	1,527
Hamap	472

Supplementary Methods

Genome Sequencing

Genomic DNA (2 µg) was sheared on a Covaris™ E220 and size selected on 2% agarose gel to obtain two insert sizes of 480-770 bp and 220-430 bp. The size selected DNA was end-repaired, adenylated, and ligated to Illumina specific indexed paired-end adaptors. Each library was run multiplexed on the GAIIx platform in 2x151 bp read length runs according to standard Illumina operation procedures. Primary data analysis was carried out with the standard Illumina pipeline. A total of 284 million paired end reads (> 85 Gb of raw sequence or 120x coverage) were produced.

Two mate pair (MP) libraries, with 3 and 5 kb fragment sizes, were constructed according to a modified Illumina protocol incorporating a biotinylated 454 linker at the junction. The resulting libraries were run on the HiSeq2000 platform in 2x101 bp read length runs as above. In total, 145 million PE reads (29 Gb, 41x GeC) and 164 million PE reads (33 Gb, 47x GeC) of raw sequence were produced, respectively. Post-processing of sequence reads involved trimming of the linker sequence. Only pairs for which at least one mate was trimmed (i.e. contained the linker and was thus a true MP and not PE contamination) were kept for scaffolding.

A fosmid library of 140,000 clones was constructed (CIB-CSIC) in the pNGS vector (Lucigen Corp.). The DNA was processed for end-sequencing (4-cutter digest, intramolecular ligation, PCR amplification of truncated insert including standard Illumina adaptors) according to the Lucigen protocol and the resulting library was run on the HiSeq2000 in PE mode, 2*101+7 bp, in one sequencing lane following Illumina instructions for the custom recipe with 4 initial dark cycles in order to overcome possible sequencing errors due to the presence of leftover of the restriction site situated after the Illumina sequencing primer position. Primary data analysis was carried out with the standard Illumina pipeline. A total of 60 million paired end reads (>12 Gb raw data) were produced.

To estimate the genome size we performed an analysis of the kmer content of the genome. Using the software Jellyfish v1.1.10[1], 17mers were extracted from the WGS PE reads and unique kmers were counted and plotted according to kmer depth (multiplicity).

Genome Assembly

Paired end reads were first filtered for contaminating sequences (phiX, *Escherichia coli* and other vector sequences) using GEM[2] with –m 0.02 (2% mismatches). Then, reads were assembled into unitgs using ABySS v1.3.5[3] with parameters: -s 300 -n 8 -k 96 -q 15. The unitigs were removed of contaminating sequences again (using BLAST+[4] and custom scripts), the ends trimmed by 50 bp and then subjected to a misassembly detection routine that detects potential misassemblies by inconsistency with the 200 bp and 500 bp PE reads. Inconsistent segments were removed from the contigs, leaving only consistent contigs, which were then scaffolded with SSPACE[5] as follows. The 200 and 500 bp fragment size PE libraries were trimmed to 75 bp and mapped with GEM with parameters: –m 0.04 --unique-mapping. The resulting mappings were converted to tab format files for input to SSPACE, which was run with the parameters: -x 0 -z 0 -k 5 -a 0.7 -n 15 -T 1 -p 1. The library insert sizes provided as parameters to SSPACE were 215 and 480 +/- 33%. The scaffolded assembly was gap-filled using GapFiller[6] with the parameters: -m 30 -o 2 -r 0.6 -n 10 -d 100 -t 15 -g 0 -T 8 -i 5. The assembly was then scaffolded with the two mate pair libraries using ABySS v1.3.4 with parameters: -n 5 -s 200 -N 10 -S 200-2000 -k 96 -l 36 -q 10. Again the assembly was gap-filled as above but with the addition of the two mate pair libraries 3 and 5 kb +/- 33% in RF orientation. Scaffolds were then broken at gaps greater than 6 kb in length (resulting from scaffolding with ABySS) and then the fosmid end library was used to do a final scaffolding using SSPACE with the same settings as before.

Comparative mapping

Genomic sequences containing the microsatellite/SNP loci used for linkage mapping[7] were searched against the turbot genome using the BLAST algorithm and the best hit with E-value <1e-20 was retained. This approach enabled us to anchor a set of scaffolds covering a large fraction of the genome to the turbot map. This correspondence was used as a reference to refine the relationship between physical and genetic map in turbot when a scaffold (in particular the larger ones) matched to more than one LG. The relationship between genetic and physic maps was drawn with MAPCHART 2.2[8].

Then, to identify syntenic patterns with closely related species within Percomorpha, the sequences of orthologous genes in these anchored scaffolds were compared by NCBI-BLAST with updated versions of model fish genomes downloaded from ftp://ftp.ensembl.org: *T. nigroviridis v.*8.61, *O. latipes* v.1.61 and *C. semilaevis*. BLAST[9] searching was performed by using an E-value threshold <1e-05[10,11]. Orthologous relationships among species and paralagous relationships within the turbot genome were represented with circle diagrams using CIRCOS[12].

Unanchored scaffolds of length > 150 kb covering altogether more than 96% of the turbot genome were predictively assigned to turbot LGs by doing a sequence search (BLAST) of the > 10,000 orthologous genes identified against *C. semilaevis, T. nigroviridis* and *O. latipes* proteomes and integrating these results with the already established collinearity between turbot and fish genomes. Additionally, we performed a sequence search using BLAST of the nucleotide sequence of the same turbot scaffolds against *G. aculeatus* genome and retained only single 1:1 matches for the same analysis. In this case, we used DNA sequence due to the poorer annotation of the stickleback genome. Those scaffolds matching in the same homologous position with at least three reference species were predictively assigned to turbot LGs.

Repetitive elements

Low complexity sequences and short repetitive motifs were analyzed using DUST v.1[13] and TRF v.4.07b[14]. RepeatMasker v.4.0.5[15] was used to screen for interspersed repeats, using the RepBase database (v.20.40 Apr 2015; available from the Genetic Information Research Institute)[16]. The program was run using the WU-BLAST[17] engine with default settings and a custom library including repetitive sequences from vertebrate species only. The genome sequence was further analyzed with the *blastn* and *tblastn* programs in WU-BLAST. The whole RepBase database was used as a query with *blastn* and all hits with higher than 70% nucleotide sequence homology with any query sequence was masked. In a second run, *tblastn* was used to search the amino acid sequences of all ORFs defined in RepBase elements against the masked genome. The combined output of these two programs was processed to extract the nucleotide sequences of the putatively full-length insertions (>80% length of the closest canonical query sequence), as described elsewhere[18]: i) the maximum number of unique hits per element were identified; ii) hits of the same element in a single contig were joined to build the longest chain, using the Chao and Miller algorithm[19]; iii) all chains were sorted by contig, direction, insertion point and score, and pooled in a single file. All chains embedded in other elements and the region of the chain with lower score where two chains overlapped were removed, thus leaving only the best unique hit per subject point; and iv) TE matches with > 70% of similarity over 100 bp were filtered and retained. The MITE-Hunter program[20] was used to identify putative miniature inverted repeat elements (MITEs). TE-derived sequences were named according to the RepBase element sequence and grouped following a standard TE classification proposed by Wicker *et al.*[21]. All sequences were edited with Bioedit v7.0.4.1[22] and aligned with MUSCLE[23], using -600 as gap penalty. The average number of substitutions per site between sequences was estimated using the Tamura-Nei model, assuming a gamma distribution of the rate of variation among sites (shape parameter = 1), with the aid of MEGA v6.06[24].

Protein-coding gene annotation

Transcript and protein alignment

Transcripts for assembly with PASA[25] were obtained as follows: first, reads from two 454 Roche rRNA-seq studies[26,27] were aligned to the final *S. maximus* assembly, "sm5", with GEM[2]. Transcript models were subsequently generated using the standard Cufflinks[28] pipeline and then added to the PASA database. In addition, 15,559 turbot ESTs[27,29], 404 CDS (July 24, 2013) and 53,749 mRNAs (Jan 29, 2014) present in NCBI were also added to PASA using GMAP[30] as the alignment engine. All of the above transcript alignments, in total 1,947,260, were then assembled by PASA, resulting in 105,044 PASA assembled transcripts.

We aligned Percomorpha proteins present in Uniprot to the turbot genome with SPALN[31], resulting in 1,857,695 CDS alignments.

Ab initio gene predictions

Ab initio gene predictions were performed on the sm5 assembly, which was masked for repeats by RepeatMasker[15] using the custom repeat library that we constructed. Low complexity repeats were left unmasked for this purpose.

GeneID[32] *ab initio* gene predictions were obtained by running GeneID v1.4 with the parameter file specific for the *Tetraodon* genus. *S. maximus* protein-coding gene annotations were also obtained using the gene prediction tool Augustus[33] v2.5.5. For this purpose we used the program's pre-existing *Homo sapiens* parameter file. GeneMark-ES[34] v2.3e gene predictions were obtained using its self-training mode.

Hence, GeneID, Augustus and GeneMark were subsequently used to predict genes on the repeat-masked sm5 assembly of the turbot genome made up of 16,463 scaffolds. The number of predicted gene models ranged from 28,826 (Augustus) to 187,934 (GeneMark), while GeneID predicted 39,351 genes.

Generation of consensus gene models

Evidence Modeler (EVM r2012-06-25)[35] was used to obtain consensus coding sequence (CDS) models using three main sources of evidence: aligned transcripts, aligned proteins, and gene predictions. EVM was run with six different sets of weights and the resulting consensus models with the best specificity and sensitivity as determined by intersection (BEDTools[36] intersect) with the transcript mappings were chosen for the final annotation.

The EVM models were cleaned of transposon sequence by using BLAST[9] to search the gene models produced by EVM against the RepeatMasker database of proteins encoded by transposable elements (TEs). All of the gene models that had a full-length hit against a repeat were discarded from the annotation. Those that had only a partial match to a TE were kept but modified to remove the sequence corresponding to the transposable element.

The consensus CDS models were then updated with UTRs and alternative exons through two rounds of PASA's annotation updates. A final round of quality control was performed, fixing reading frames and intron phases, and then the resulting transcripts were clustered into genes using shared splice sites or significant sequence overlap as criteria for designation as the same gene. Systematic identifiers with the prefix "SMAX5B" were assigned to the genes, transcripts and protein products derived from them.

Support by source of evidence at the gene and exon level was determined *a posteriori* using BEDTools[36] intersect and multiinter programs.

Proteome functional annotation

We used InterPro[37], KEGG[38] and Blast2GO[39] databases for functional annotation. InterProScan v.5[40] was used to scan through all available InterPro databases, including the most important ones - PANTHER[41], Pfam[42], TIGRFAM[43], HAMAP[44] and SUPERFAMILY[45]. BLAST search against NCBI non-redundant (NR) collection of protein sequences (release 2014-02) was used as input to the local Blast2GO software p2gpipe version 2.5.0. KEGG orthology (KO) groups were assigned by KEGG

Automatic Annotation Server (KAAS)[46] using bi-directional best hit (BBH) method against a representative gene set from 28 different species, including *D. rerio.* KO identifiers were then used to retrieve using the KEGG REST-based API service the KEGG relevant functional annotation, such as metabolic pathways and external database references. Distribution of GO terms grouped by the different functional categories was done with CateGOrizer[47] by using GOSlim without top-level categories

Phylogenomics

Phylome reconstruction

Turbot phylome was reconstructed using the phylomeDB pipeline[48]. For each turbot gene, a search was performed against a database containing the proteomes of the 17 selected species (Table MS5). We used an e-value threshold <1e-05 and a continuous overlap of 50% over the query sequence for the detection of homologues, and limited the number of hits to the closest 150 homologues per gene. Multiple sequence alignments of homologous sequences were built using three different aligners, which were used in forward and reverse orientations (MUSCLE[23], MAFFT[49] and KALIGN[50]. The resulting six alignments were combined using M-COFFEE[51], and then, trimAl v1.4[52] was used to trim the alignment (consistency cut-off of 0.16667 and -gt >0.1). Subsequently, trees were constructed using PhyML v3[53] using the best fitting model, four rate categories with rates and fraction of invariant sites estimated from the data. Branch support was analyzed using an aLRT (approximate likelihood ratio test) non-parametric test based on a chi-square distribution.

Prediction of orthology and paralogy relationships

Paralogy and orthology predictions were analyzed based on phylogenetic evidence from the turbot phylome. ETE[54] was used to infer gene duplication and speciation events with a species overlap approach (species overlap score of 0). Orthologous genes are those who the last common ancestor is represented by a speciation event, and paralogous genes are those that diverge from duplication events[55]. All trees, alignments, and information about orthology and paralogy relationships are available in phylomeDB[56] with the PhylomeID code 18.

Gene duplications

The turbot phylome was analyzed to detect genes that had undergone duplications in lineages leading to this species using a previously-described algorithm of duplication detection and dating[57]. Gene enrichment was analyzed using FatiGO[58] by comparing annotations of the proteins involved in a duplication at a given age against all the others encoded in the genome.

Species tree reconstruction

A total of 389 genes with one-to-one orthologues in each studied species were selected and their trimmed alignments concatenated. We then used RAxML v7.2.6, model Protgammalg[59] to derive the species tree. Bootstrap supports were calculated by creating 100 alignments using Phylip's

SeqBoot[60]. Finally, we reconstructed a super-tree from all single gene trees in the turbot phylome using a gene tree parsimony strategy as implemented in duptree[61].

Adaptation to benthic life

Functional enrichment analysis of paralogous genes were carried out using the KOBAS web server (http://kobas.cbi.pku.edu.cn/help.do). For phylogenetic analysis of selected genes, protein sequences from sequenced fishes and those corresponding to *C. semilaevis* were downloaded from ENSEMBL and the NCBI databases respectively. ClustalW[62] using Gonnet's protein weight matrix was used to produce alignments which were then visually inspected and filtered using Guidance2[63]. Alignments were trimmed using trimAL[52] to remove poorly aligned regions, and phylogenetic reconstruction was performed by the neighbor joining method using the JTT aminoacid substitution model in MEGA6[24]. Statistical support of the trees was obtained using 1,000 bootstrap replicates. The ratio Ka/Ks was calculated following the Nei-Gojobori model[64]. All ambiguous positions were removed for each sequence pair. Evolutionary analyses were conducted in MEGA6[24].

Genetic architecture of growth, resistance to diseases and sex determination

Mining analysis of the turbot genome was performed to characterize previously detected QTL regions and to identify candidate genes influencing biological pathways related to sex, growth and disease resistance traits. A set of selected QTL markers related to these three features were located in the turbot genome using BLAST[9]. Candidate genes for growth-, sex- and disease resistance-related traits in fish and vertebrates were selected based on previous reports and our own data and then mapped to the turbot genome taking QTL markers as a reference (Supplementary Tables 12 and 13). For this, we used the relationship between physical and genetic maps established in our work (Supplementary Table 3; Supplementary Fig. 2). Gene lists were extracted from conservative 2-Mb windows surrounding each selected QTL for all traits (closest associated marker position ± 1 Mb), assuming an average genomic relationship of 0.5 Mb/cM[11] (Supplementary Table 14). To identify suggestive candidate genes and pathways within the extracted gene lists, we performed Gene ontology (GO)[65] enrichment analysis using BLAST2GO[66] and KEGG[38] pathway enrichment using KOBAS[67], against the turbot proteome (22,751 genes) as background (Supplementary Table 15). Enrichment probability values were adjusted for multiple testing (False Discovery Rate (FDR)-corrected P-values < 0.05). All analyses were focused on major QTL, either associated with sex determination, growth or disease resistance, but also on overlapping QTL for different traits[68] (Supplementary Table 16). A large set of QTL markers related to sex determination-SD (4), growth traits (12 for body weight-BW, 9 for length-L and 6 for Fulton's factor-FK), and disease resistance (7 for *Aeromonas salmonicida*-AS, 19 for *Philasterides dicentrarchi*-PD and 16 for Viral Hemorrhagic Septicaemia Virus-VHSV) were physically mapped to scaffolds of the turbot genome and mined. Selected candidate genes identified across QTL regions and traits were represented into the turbot genetic map[7] using Mapchart 2.2[8].

Supplementary Results Genetic architecture of sex determination, growth and resitance to diseases

The highest concentration of candidate genes related to sex determination was detected on LG5 (12 genes), LG6 (5), LG8+18 (7) and LG21+24 (9). Additionally, 12 genes were found on LG1, not previously related to SD. The mining strategy around SD-QTL revealed 23 additional genes involved in sex differentiation: *lhx9*, *bcar3* and *dmrt2b* on LG5; *cyp19a1* and *cyp11a* on LG6; *ar*, *lhx1* and *foxo1* at LG8; and *ryr2a*, *sox17*, *sox8a*, *sox9a* and *rara* on LG21+24 (Supplementary Table S15A).

Genetic factors (Supplementary Table S15B) and functional enrichment related to regulation of muscle development and growth (Supplementary Table S16B) were found at specific QTL. Different pathways related to L (mucin O-glycan biosynthesis) and BW (arachidonic acid metabolism and taste transduction) were identified at different QTL supporting distinct genetic mechanisms underlying growth traits. The most significant pathway involved in muscle differentiation and lipid metabolism was extracellular matrix communication (ECM)-receptor interaction (*gh1*, *lamb1*, *itga11*), associated to a L-QTL (LG6). ECM has been associated with growth in fish and other vertebrates[69,70]. Pathway enrichment for taste transduction and arachidonic acid metabolism was detected within a BW-QTL (LG11), which includes candidate genes (*tas1r3* and *gpx1*) associated with lipid metabolism and growth effects of aquaculture diets[71]. Non-homologous end-joining and mucin type O-glycan biosynthesis pathways found in L-QTL (LG17 and LG20, respectively), also pinpoint candidate genes (*fen1*, *galnt3*, *galnt5*, *galnt13*) previously associated to differential growth in vertebrates[72,73].

Relevant immune genes were identified within VHS-, AS-, and PD-QTL (Supplementary Table S15C). Virus defense and clearance related genes were detected within VHS-QTL: i) genes implicated in T-cell proliferation, differentiation, maturation or activation on several LGs (*nlrc3*, *malt1*, *vav*, *nfkbid*, *irf4*); ii) genes involved in the blood coagulation cascade (*thpo*, *lrp8*, *f3* at LG1; *clec3b*, *thbs1*, *plgrkt*, *ptgds*, *plek* on LG2; *pip5k1c*, *bsg* at LG5; *vwf, cd9*, *calu*, *slc7a5*, *ranbp10* at LG6; and *plscr1*, *serpinb6* on LG17) likely related to the important hemorrhagic activity of this virus; and iii) genes related to iron homeostasis and scavenging[74], like some transferrin related genes (*tf*, t*frc*; LG1) and hepcidin (*hamp*; LG2). In fact, hepcidin was previously associated with resistance to VHSV in turbot families[75]. Typical antibacterial and bacterial recognition genes were found around AS-QTL: peptidoglycan recognition protein 1 (*pglyrp1*), g-type lysozyme (*lyg1*) and macrophage mannose receptor 1-like (*mrc1*) (tightly linked on LG9). Finally, important immune genes were also detected at PD-QTL: interleukin-17 (*il17f*; LG16) and its receptor (*il17re*; LG23); galectin-8 (*lgals8*; LG3), also reported as candidate for ISAV-resistance in *Salmo salar*[76]; perforin-1 precursor (*prf1*; LG9), related to a broad antimicrobial spectrum[77]; and two toll-like receptors (TLR) (*tlr2* and *tlr3*; LG9). The activity of TLRs in response to parasitic infections has been widely documented[78].

Finally, we detected genes and functions underlying genomic regions associated to different traits in turbot (Supplementary Table S17; Fig. 5). Among them, we identified in an overlapping BW- and VHSV-QTL region (LG1) genes associated with growth like *myod1*, a gene which plays a central role in the development of the skeletal muscle in fish and vertebrates[79] and recently associated with meat quality in rainbow trout[80,81], and genes involved in the TNF signaling pathway (*tab2*) which are associated with immune response in fish[82]. Also, the BW- and BL-QTL on LG5 include a sex-associated marker (ScmM1)[83] linked to *tgfbr3* and *sox14*, genes related to reproduction and cell proliferation in fish[84]. The relationship between sex and growth in fish has been widely documented and it is of special interest in turbot considering its sexual dimorphism in growth[85].

Supplementary References

1. Marçais, G. and Kingsford, C. 2011, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers, *Bioinformatics*, 27, 764-770.

2. Marco-Sola, S., Sammeth, M., Guigo, R. and Ribeca, P. 2012, The GEM mapper: fast, accurate and versatile alignment by filtration, *Nat. Methods*, 9, 1115-1118.

3. Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones ,S. J., Birol, I. 2009, ABySS: a parallel assembler for short read sequence data, *Genome Res.*, 19, 1117-1123.

4. Camacho, C., Coulouris, G., Avagyan, V., et al. 2009, BLAST+: architecture and applications, *BMC Bioinformatics*, 10, 421.

5. Boetzer, M., Henkel, C. V., Jansen, H. J. Butler, D. and Pirovano, W. 2011, Scaffolding pre-assembled contigs using SSPACE, *Bioinformatics*, 27, 578–579.

6. Boetzer, M. and Pirovano, W. 2012, Toward almost closed genomes with GapFiller, *Genome Biol.*, 13, R56.

7. Hermida, M., Bouza, C., Fernández, C., et al. 2013, Compilation of mapping resources in turbot (*Scophthalmus maximus*): A new integrated consensus genetic map, *Aquaculture*, 414-415, 19-25.

8. Voorrips, R. E. 2002, MapChart: software for the graphical presentation of linkage maps and QTLs, *J. Hered.*, 93, 77–78.

9. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 1990, Basic local alignment search tool, *J. Mol. Biol.*, 215, 403-410.

10. Martínez, P., Bouza, C., Hermida, M., et al. 2009, Identification of the major sex-determining region of turbot (*Scophthalmus maximus*), *Genetics*, 183, 1443–1452.

11. Bouza, C., Hermida, M., Pardo, B. G., et al. 2012, An Expressed Sequence Tag (EST)-enriched genetic map of turbot (*Scophthalmus maximus*): a useful framework for comparative genomics across model and farmed teleosts, *BMC Genet.*, 13, 54.

12. Krzywinski, M., Schein, J., Birol, I., et al. 2009, Circos: an information aesthetic for comparative genomics, *Genome Res.*, 19,1639-1645.

13. Morgulis, A., Gertz, E. M., Schäffer, A. A. and Agarwala, R. 2006, A fast and symmetric DUST implementation to mask low-complexity DNA sequences, *J. Comput. Biol.*, 13, 1028-1040.

14. Benson, G. 1999, Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Res.*, 27, 573-580.

15. Smit, A. F. A., Hubley, R. and Green, P. 2013-2015, RepeatMasker Open-4.0, http://www.repeatmasker.org.

16. Jurka, J, Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J. 2005, Repbase Update, a database of eukaryotic repetitive elements, *Cytogenet. Genome Res.*, 110, 462-467.

17. Gish, W. 1996-2003, http://blast.wustl.edu.

18. Bartolomé, C., Bello, X. and Maside, X. 2009, Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes, *Genome Biol.*, 10, R22.

19. Chao, K. M., Zhang, J., Ostell, J. and Miller, W. 1995, A local alignment tool for very long DNA sequences, *Comput. Appl. Biosci.*, 11, 147-153.

20. Han, Y. and Wessler, S. R. 2010, MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences, *Nucleic Acids Res.*, 38**,** e199.

21. Wicker, T., Sabot, F., Hua-Van, A., et al. 2007, A unified classification system for eukaryotic transposable elements, *Nat. Rev. Genet.*, 8, 973-982.

22. Hall, T. A. 1999, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucl. Acids Symp. Ser.*, 41, 95-98.

23.Edgar, R. C. 2004, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res*., 32, 1792-1797.

24. Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. 2013, MEGA6: Molecular evolutionary genetics analysis version 6.0, *Mol. Biol. Evol.*, 30, 2725-2729.

25. Haas, B. J., Delcher, A. L., Mount, S. M., et al. 2003, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies, *Nucleic Acids Res.*, 31, 5654-5666.

26. Pereiro, P., Balseiro, P., Romero, A., et al. 2012, High-Throughput sequence analysis of turbot (*Scophthalmus maximus*) transcriptome using 454-pyrosequencing for the discovery of antiviral immune genes, *PLoS ONE,* 7, e35369.

27. Ribas, L., Pardo, B. G., Fernández, C., et al. 2013, A combined strategy involving Sanger and 454 pyrosequencing increases genomic resources to aid in the management of reproduction, disease control and genetic selection in the turbot (*Scophthalmus maximus*), *BMC Genomics*, 14, 180.

28. Trapnell, C. 2010, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, *Nat. Biotechnol.*, 28, 511-515.

29. Pardo, B. G., Fernández, C., Millán, A., et al. 2008, Expressed sequence tags (ESTs) from immune tissues of turbot (*Scophthalmus maximus*) challenged with pathogens, *BMC Vet. Res.*, 4, 37.

30. Wu, T. and Watanabe, C. 2005, GMAP: a genomic mapping and alignment program for mRNA and EST sequences, *Bioinformatics*, 21, 1859-1875.

31. Iwata, H. and Gotoh, O. 2012, Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features, *Nucleic Acids Res.*, 40, e161.

32. Blanco, E., Parra, G. and Guigó, R. 2007, Using geneid to identify genes, *Curr. Protoc. Bioinformatics*, 18, 4.3:4.3.1–4.3.28.

33. Stanke, M. and Waack, S. 2003, Gene prediction with a hidden Markov model and a new intron submodel, *Bioinformatics*, 19, ii215-ii225.

34. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O. and Borodovsky, M. 2008, Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training, *Genome Res.*, 18, 1979-1990.

35. Haas, B. J., Salzberg, S. L., Zhu, W., et al. 2008, Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments, *Genome Biol.*, 9, R7.

36. Quinlan, A. R. and Hall, I. M. 2010, BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics*, 26, 841-842.

37. Hunter, S., Jones, P., Mitchell, A., et al. 2011, InterPro in 2011: new developments in the family and domain prediction database, *Nucleic Acids Res.*, 40, D306–D312.

38. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M. 2012, KEGG for integration and interpretation of large-scale molecular datasets, *Nucleic Acids Res.*, 40, D109-D114.

39. Götz, S., García-Gómez, J. M., Terol, J., et al. 2008, High-throughput functional annotation and data mining with the Blast2GO suite, *Nucleic Acids Res.*, 36, 3420-3435.

40. Zdobnov, E. M. and Apweiler, R. 2001, InterProScan an integration platform for the signature-recognition methods in InterPro, *Bioinformatics*, 17, 847-848.

41. Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S., Thomas, P. D. 2010, PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium, *Nucleic Acids Res.*, 38, D204-210.

42. Finn, R. D., Bateman, A., Clements, J., et al. 2014, Pfam: the protein families database, *Nucleic Acids Res.*, 42, D222-D230.

43. Selengut, J. D., Haft, D. H., Davidsen, T., et al. 2007, TIGRFAMs and genome properties: tools for the assignment of molecular function and biological process in prokaryotic genomes, *Nucleic Acids Res.*, 35, D260–D264.

44. Lima, T., Auchincloss, A. H., Coudert, E., et al. 2009, HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot, *Nucleic Acids Res.*, 37, D471–D478.

45. de Lima Morais, D. A., Fang, H., Rackham, O. J., et al. 2011, SUPERFAMILY 1.75 including a domain-centric gene ontology method, *Nucleic Acids Res.*, 39, D427–D434.

46. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. and Kanehisa, M. 2007, KAAS: an automatic genome annotation and pathway reconstruction server, *Nucleic Acids Res.*, 35, W182-W185.

47. Hu, Z. L., Bao, J. and Reecy, J. M. 2008, CateGOrizer: A web-based program to batch analyze Gene Ontology classification categories, *Online J. Bioinform.*, 9, 108-112.

48. Huerta-Cepas, J., Capella-Gutierrez, S., Pryszcz, L. P., et al. 2011, PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions, *Nucleic Acids Res.*, 39, D556-560.

49. Katoh, K., Kuma, K., Toh, H. and Miyata, T. 2005, MAFFT version 5: improvement in accuracy of multiple sequence alignment, *Nucleic Acids Res.*, 33, 511–518.

50. Lassmann, T. and Sonnhammer, E. L. 2005, Kalign an accurate and fast multiple sequence alignment algorithm, *BMC Bioinformatics*, 6, 298.

51. Wallace, I. M., O'Sullivan, O., Higgins, D. G. and Notredame, C. 2006, M-Coffee: combining multiple sequence alignment methods with T-Coffee, *Nucleic Acids Res.*, 34, 1692–1699.

52. Capella-Gutiérrez, S., Silla-Martínez, J. M. and Gabaldón, T. 2009, TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, *Bioinformatics*, 25, 1972-1973.

53. Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. 2010, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, *Syst. Biol.*, 59, 307–321.

54. Huerta-Cepas, J., Dopazo, J. and Gabaldón, T. 2010, ETE: a python Environment for Tree Exploration,*BMC Bioinformatics*, 11, 24.

55. Gabaldón, T. 2008, Large-scale assignment of orthology: back to phylogenetics?, *Genome Biol.*, 9, 235.

56. Huerta-Cepas, J., Capella-Gutiérrez, S., Pryszcz, L. P., Marcet-Houben, M. and Gabaldón, T. 2014, PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome, *Nucleic Acids Res.*, 42, D897-D902.

57. Huerta-Cepas, J. and Gabaldón, T. 2011, Assigning duplication events to relative temporal scales in genome-wide studies, *Bioinformatics*, 27, 38-45.

58. Al-Shahrour, F. 2007, FatiGO+: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments, *Nucleic Acids Res.*, 35, W91-W96.

59. Stamatakis, A. 2006, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models, *Bioinformatics*, 22, 2688-2690.

60. Felsenstein, J. 2005, SEQBOOT—bootstrap, jackknife or permutation resampling of molecular sequence, restriction site, gene frequency or character data.

61. Wehe, A., Bansal, M. S., Burleigh, J. G. and Eulenstein, O. 2008, DupTree: a program for large-scale phylogenetic analyses using gene tree parsimony, *Bioinformatics,* 24, 1540–1541.

62. Larkin, M. A., Blackshields, G., Brown, N. P., et al. 2007, ClustalW and ClustalX version 2, *Bioinformatics*, 23, 2947-2948.

63. Sela, I., Ashkenazy, H., Katoh, K. and Pupko, T. 2015, GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters, *Nucleic Acids Res.*, 43, W7-W14.

64. Nei, M. and Gojobori, T. 1986, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, *Mol. Biol. Evol.*, 3, 418-426.

65. Ashburner, M., Ball, C. A., Blake, J. A., et al. 2000, Gene ontology: tool for the unification of biology, *Nat. Genet.*, 25, 25-29.

66. Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., Robles, M. 2005, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics*, 21, 3674-3676.

67. Xie, C. Mao, X., Huang, J., et al. 2011, KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases, *Nucleic Acids Res.*, 39, W316-W322.

68. Rodríguez-Ramilo, S. T., De La Herrán, R., Ruiz-Rejón, C., et al. 2014, Identification of quantitative trait loci associated with resistance to viral haemorrhagic septicaemia (VHS) in turbot (*Scophthalmus maximus*): a comparison between bacterium, parasite and virus diseases, *Mar. Biotechnol.*, 16, 265-276.

69. Moghadam, H. K., Ferguson, M. M., Rexroad, C. E. 3rd, Coulibaly, I. and Danzmann, R. G. 2007, Genomic organization of the IGF1, IGF2, MYF5, MYF6, and GRF/PACAP genes across Salmoninae genera, *Anim. Genet*., 38, 527-532.

70. Lee, H. J., Jang, M., Kim, H., et al. 2013, Comparative transcriptome analysis of adipose tissues reveals that ECM-receptor interaction is involved in the depot-specific adipogenesis in cattle, *PLoS ONE*, 8, e66267.

71. Matsumoto, I., Ohmoto, M. and Abe, K. 2013, Functional diversification of taste cells in vertebrates, *Semin. Cell Dev. Biol.*, 24, 210-214.

72. Long, Y., Li, Q., Zhou, B., Song, G., Li, T. and Cui, Z. 2013, De novo assembly of mud loach (*Misgurnus anguillicaudatus*) skin transcriptome to identify putative genes involved in immunity and epidermal mucus secretion, *PLoS ONE*, 8, e569982013.

73. Doran, A. G., Berry, D. P. and Creevey, C. J. 2014, Whole genome association study identifies regions of the bovine genome and biological pathways involved in carcass trait performance in Holstein-Friesian cattle, *BMC Genomics*, 15, 837.

74. Drakesmith, H. and Prentice, A. 2008, Viral infection and iron metabolism, *Nat. Rev. Microbiol.*, 6, 541-552.

75. Díaz-Rosales, P., Romero, A., Balseiro, P., Dios, S., Novoa, B. and Figueras, A. 2012, Microarray-based identification of differentially expressed genes in families of turbot (*Scophthalmus maximus*) after infection with viral haemorrhagic septicaemia virus (VHSV), *Mar. Biotechnol.*, 14, 515-529.

76. Li, J., Boroevich, K. A., Koop, B. F. and Davidson, W. S. 2011, Comparative genomics identifies candidate genes for infectious salmon anemia (ISA) resistance in Atlantic salmon (*Salmo salar*). *Mar. Biotechnol.*, 13, 232-241.

77. Trapani, J.A. and Smyth, M. J. 2002, Functional significance of the perforin/granzyme cell death pathway, *Nat. Rev. Immunol*., 2, 735-747.

78. Gundra, U. M., Mishra, B. B., Wong, K. and Teale, J. M. 2011, Increased disease severity of parasite-infected TLR2-/- mice is correlated with decreased central nervous system inflammation and reduced numbers of cells with alternatively activated macrophage phenotypes in a murine model of neurocysticercosis, *Infect. Immun.*, 79, 2586-2596.

79. Andersen, Ø., Dahle, S.W., van Nes, S.,et al. 2009, Differential spatio-temporal expression and functional diversification of the myogenic regulatory factors MyoD1 and MyoD2 in Atlantic halibut (*Hippoglossus hippoglossus*), *Comp. Biochem. Physiol. B Biochem*. *Mol. Biol*., 154, 93-101.

80. Chen, W. X., Ma, Y., and Liu, K. H. 2014, Association of MyoD1a and MyoD1b gene polymorphisms and meat quality traits in rainbow trout, *Genet. Mol. Res.*, 14, 9034-9044.

81. Johnston, I. A., Bower, N. I. and Macqueen, D. J. 2011, Growth and the regulation of myotomal muscle mass in teleost fish, *J. Exp. Biol.*, 214, 1617-1628.

82. Zhao, F., Li, Y. W., Pan, H. J., et al. 2014, TAK1-binding proteins (TAB1 and TAB2) in grass carp (*Ctenopharyngodon idella*): identification, characterization, and expression analysis after infection with *Ichthyophthirius multifiliis*, *Fish Shellfish Immunol*., 38, 389-399.

83. Viñas, A., Taboada, X., Vale, L., et al. 2012, Mapping of DNA sex-specific markers and genes related to sex differentiation in turbot (*Scophthalmus maximus*), *Mar. Biotechnol.*, 14, 655-663.

84. Mazzuchelli, J., Yang, F., Kocher, T. D. and Martins, C. 2011, Comparative cytogenetic mapping of Sox2 and Sox14 in cichlid fishes and inferences on the genomic organization of both genes in vertebrates, *Chromosome Res.*, 19, 657-667.

85. Martínez, P., Viñas, A. M., Sánchez, L., Díaz, N., Ribas, L. and Piferrer, F. 2014, Genetic architecture of sex determination in fish: Applications to sex ratio control in aquaculture, *Front. Genet.*, 5, 340.