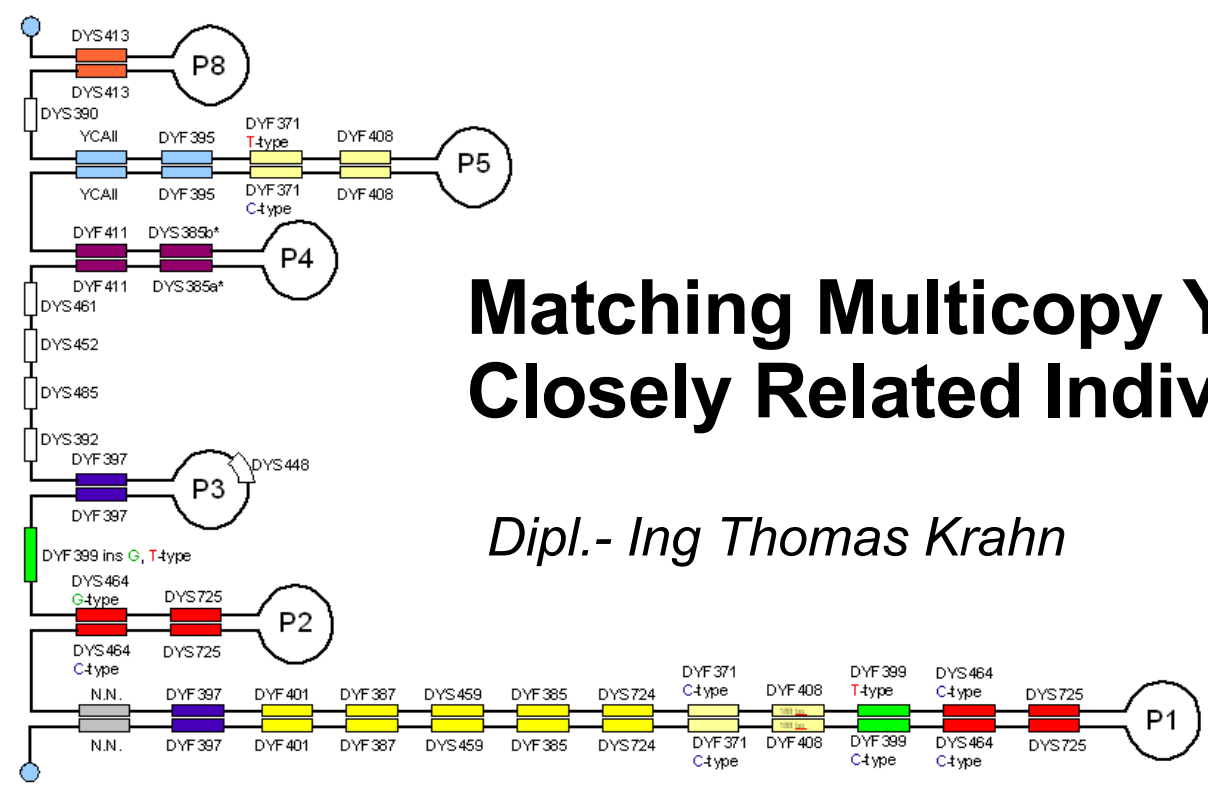


# Matching Multicopy Y-STR Markers In Closely Related Individuals

*Dipl.- Ing Thomas Krahn*



**Matching multicopy Y-STR markers in closely related individuals** - Multicopy Y-STR markers are located on the symmetrical arms of palindromes. The DNA sequences on the palindromic arms are nearly identical and therefore highly prone to intrachromosomal recombination events. The talk will demonstrate how the recombination mechanisms will affect apparent mismatches in closely related individuals and it will present new tools that can help to understand the intrachromosomal re-arrangements in the Yq11 palindromic region.

# What's a palindrome?

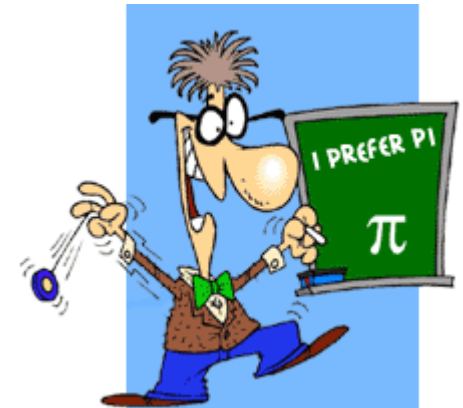
## Palindromic word phrases

### RADAR

Madam, I'm Adam.  
So many dynamos.  
Cleveland DNA: Level C  
Red rum, sir, is murder.  
Too bad, I hid a boot.  
Sat in a taxi, left Felix at Anita's  
Emil asleep, Hannah peels a lime.  
Anne, I stay a day at Sienna.  
Was it a car or a cat I saw?



A dog!  
A panic in a pagoda!



Max, I stay away at six a.m.  
Go hang a salami! I'm a lasagna hog!  
Yawn a more Roman way!

## What's a palindrome?

### Palindromic word phrases:

Forward = Backward      **RADAR**

### Nucleic acids:

Forward = Reverse Complement (Forward)

**AGCTTCTAGTCGACTAGGAAGCT**

## What's a palindrome?

### Nucleic acids:

Forward = Reverse Complement (Forward)

**AGCTTCTAGTCGACTAGGAAGCT**

Reverse Complement (**AGCTTCTAGTCGACTAGGAAGCT**)

= **AGCTTCTAGTCGACTAGGAAGCT**



## What's a palindrome?

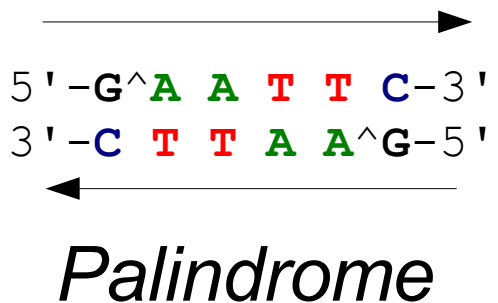
### Nucleic acids:

Forward = Reverse Complement (Forward)

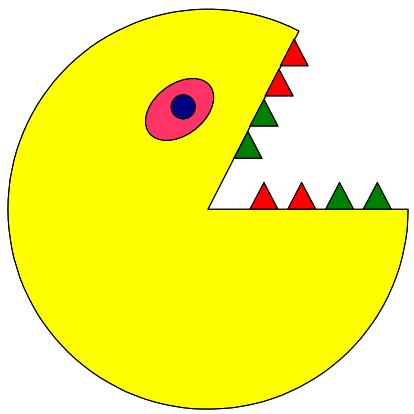


# Biological Relevance of Palindromes

Restriction enzymes cut DNA at palindromic recognition sequences



EcoRI

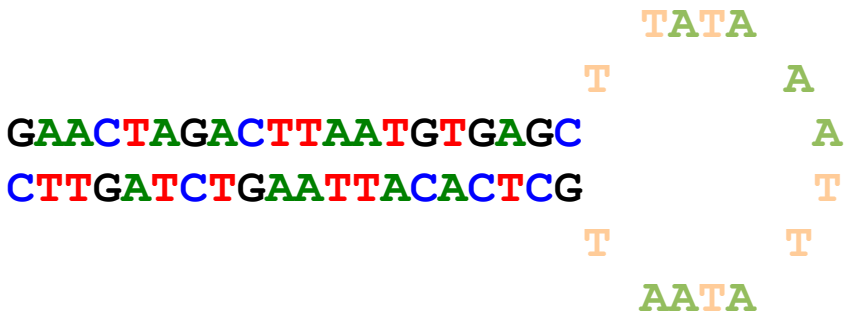
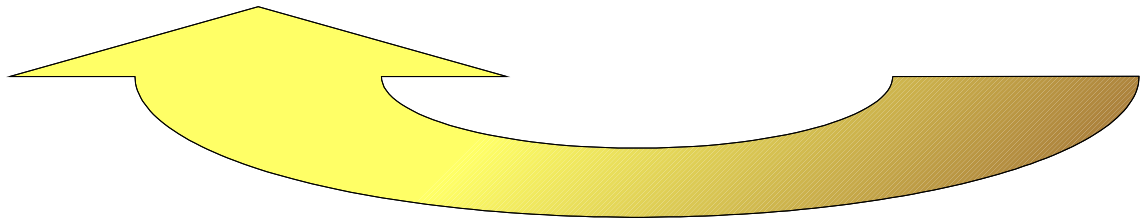


# Biological Relevance of Palindromes

## Formation of hairpins

*Single stranded DNA with partial palindromic sequence*

GAAC**TAGACTTAA**TGTGAGC**TTATAAATTATAA**TGCTCTCAC**TTAAGTCTAGTTC**

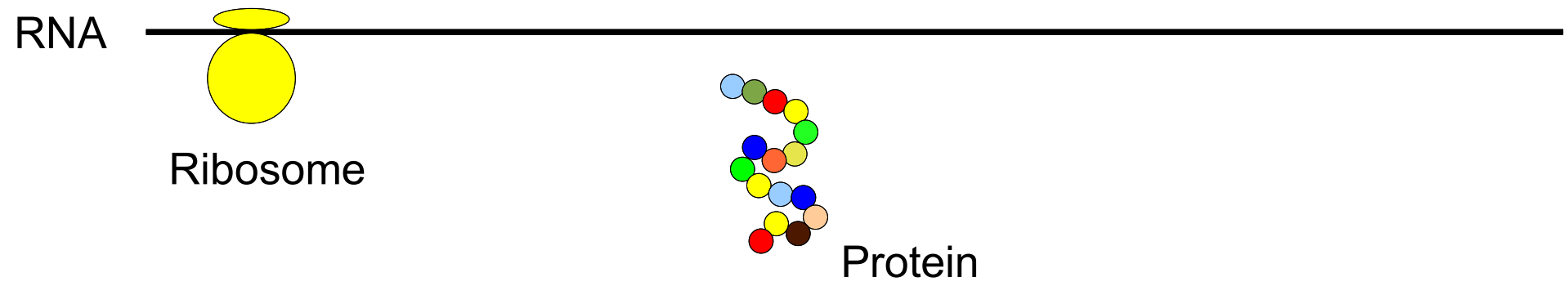


*Hairpin structure*

# Biological Relevance of Palindromes

## Hairpins regulate RNA translation

*High temperature / low ionic concentration*



*Low temperature / high ionic concentration*

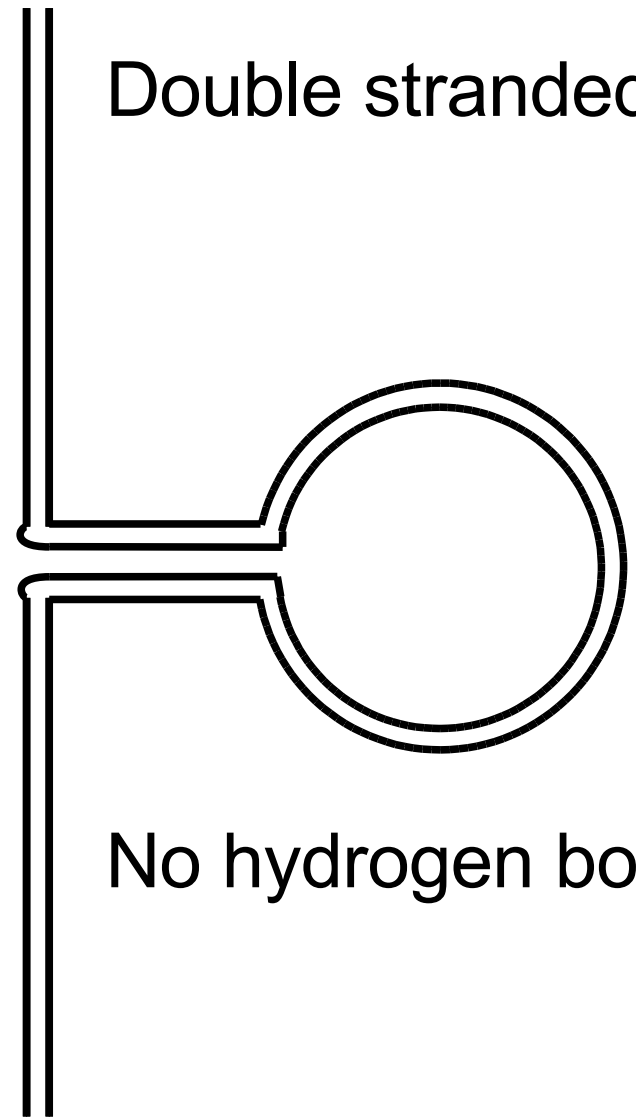




# Palindromes in the ds Human Genome



Chromosome



Double stranded DNA

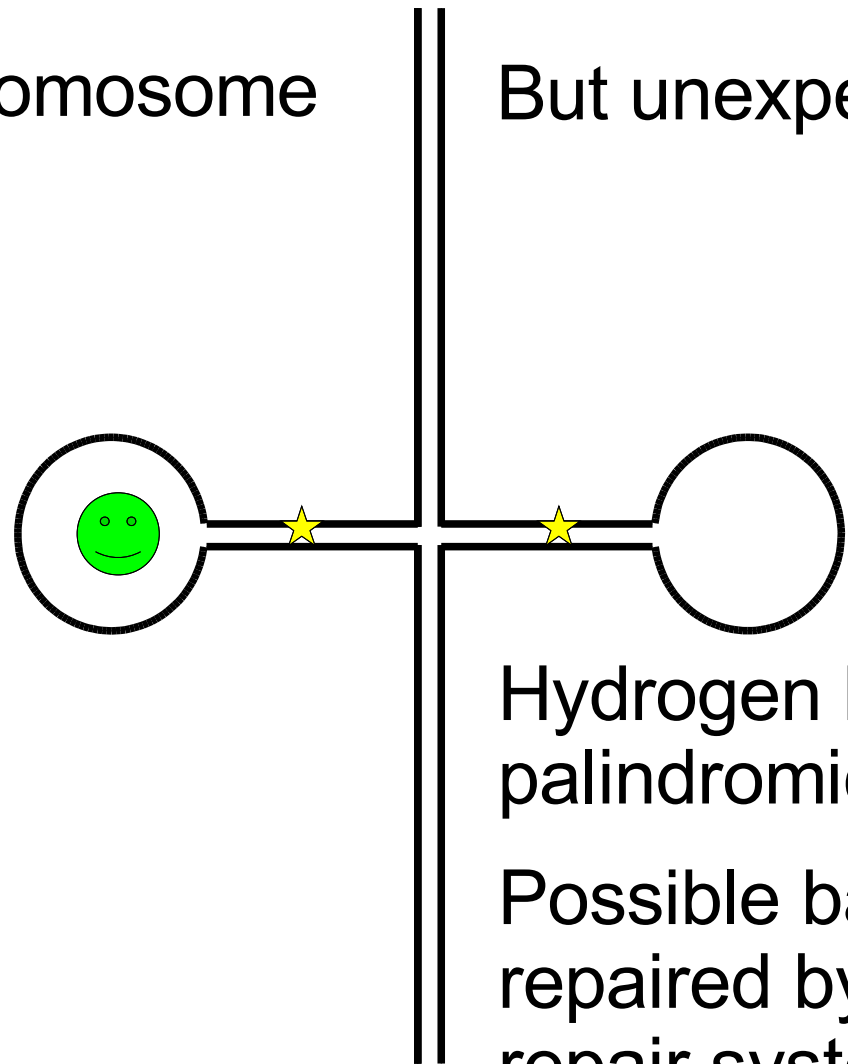
No hydrogen bond interaction

# Palindromes in the ds Human Genome



Chromosome

But unexpected things happen...



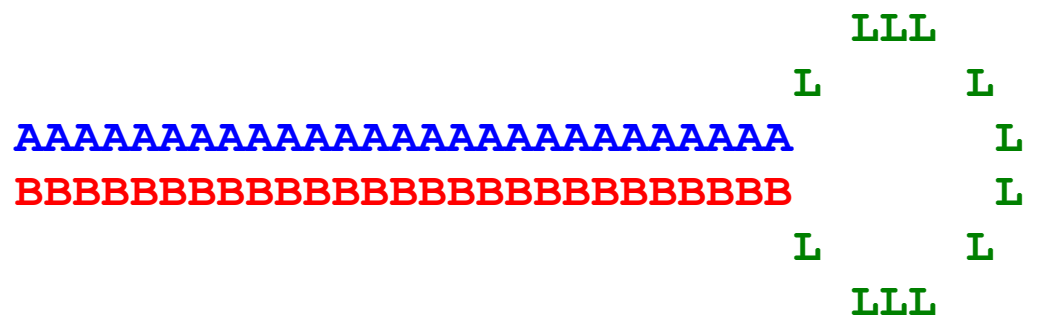
Hydrogen bonding between palindromic arms

Possible base differences will be repaired by the cell's mutation repair system

**recombinational loss of heterozygosity (recLOH)**

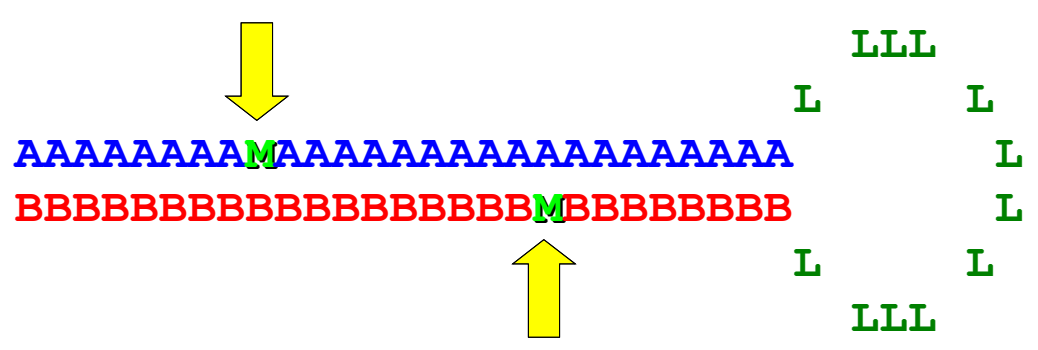
# Recombinational Loss Of Heterozygosity (recLOH)

If we have a palindromic region (that means, that the DNA sequence has a loop and the ends of the loop can be put parallel next to each other and the parallel ends have nearly exactly the same DNA sequence, in other words an inverted repeat) we can model a simple structure like this:



# Recombinational Loss Of Heterozygosity (recLOH)

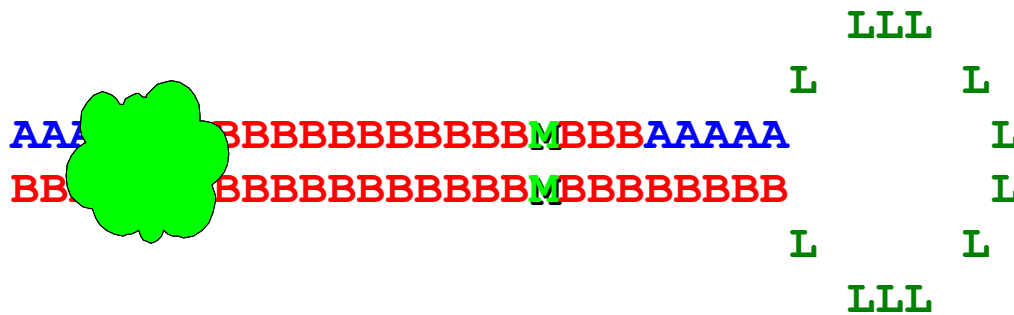
Palindromic sequences develop independently on each arm, so after some time they acquire mismatches (M) that could be STR mutations or SNPs on both arms. If it comes to a hairpin conformation, base pairing is not perfect and the cell tries to repair the apparent mutations.





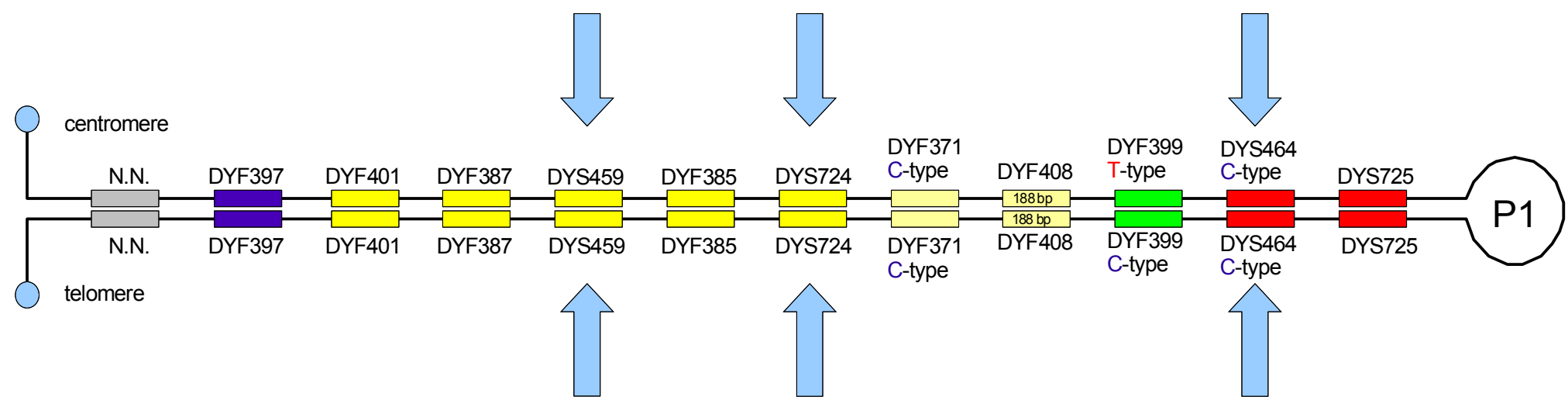
# Recombinational Loss Of Heterozygosity (recLOH)

Depending on the direction of the repair enzyme complex, some mutations get duplicated on the other arm and some mutations disappear.



# Recombinational Loss Of Heterozygosity (recLOH)

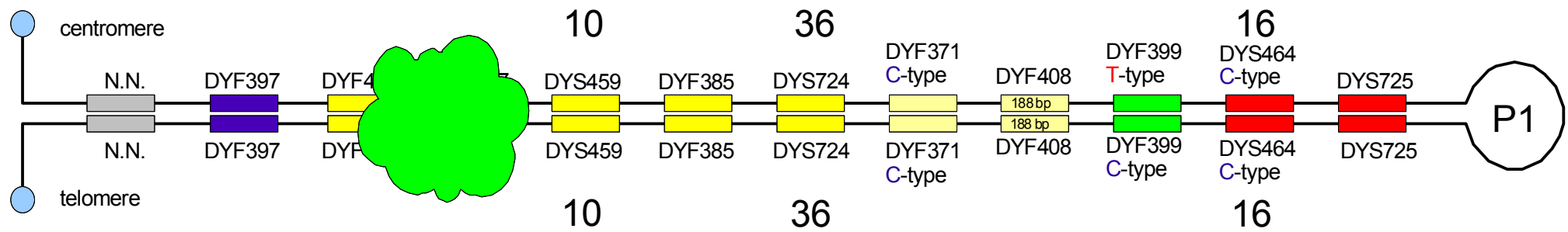
DYS459, DYS724 (CDY) and DYS464 are on the same palindrome called P1



# Recombinational Loss Of Heterozygosity (recLOH)

In this example the haplotype has different alleles at all marker pairs

After the recombination event the heterozygosity is lost => recLOH



# Typical recLOH Patterns In Genetic Genealogy

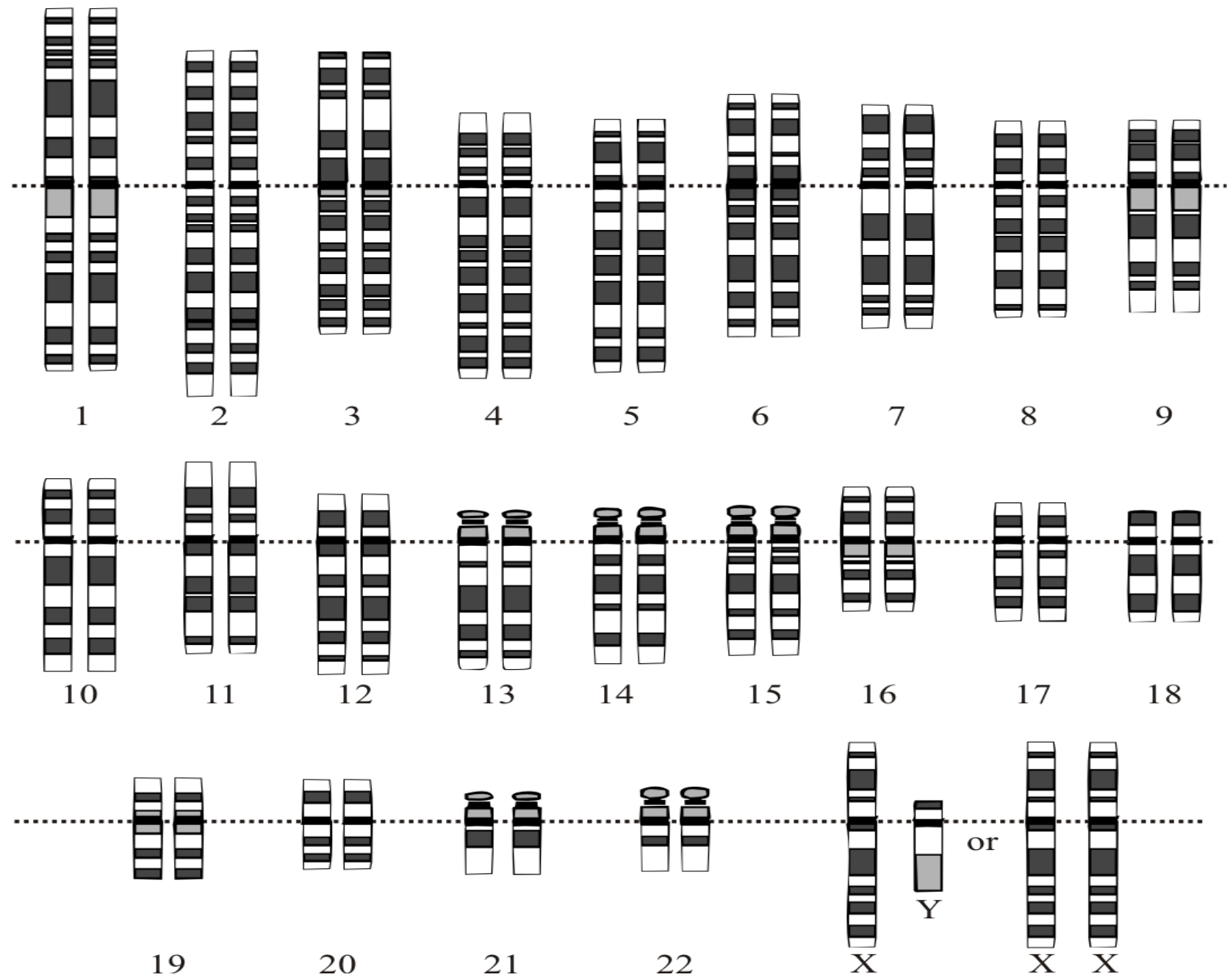
A classical DYS459, CDY, DYS464 recLOH

DYS459	DYS724 (CDY)	DYS464	Genetic distance?
9-10	37-38	14-14-16-18	
9-10	37-38	14-15-16-18	1
9-10	37-38	14-14-16-18	0
9-10	37-39	14-14-16-18	1
9-10	37-38	14-14-16-18	0
10-10	38-38	14-14-18-18	1
9-9	37-37	14-16-18-18	1

Similar recLOH events happen at all palindromic multicopy markers

DYS385	YCAII
11-14	19-23
11-14	19-23
11-14	19-19
11-11	19-23





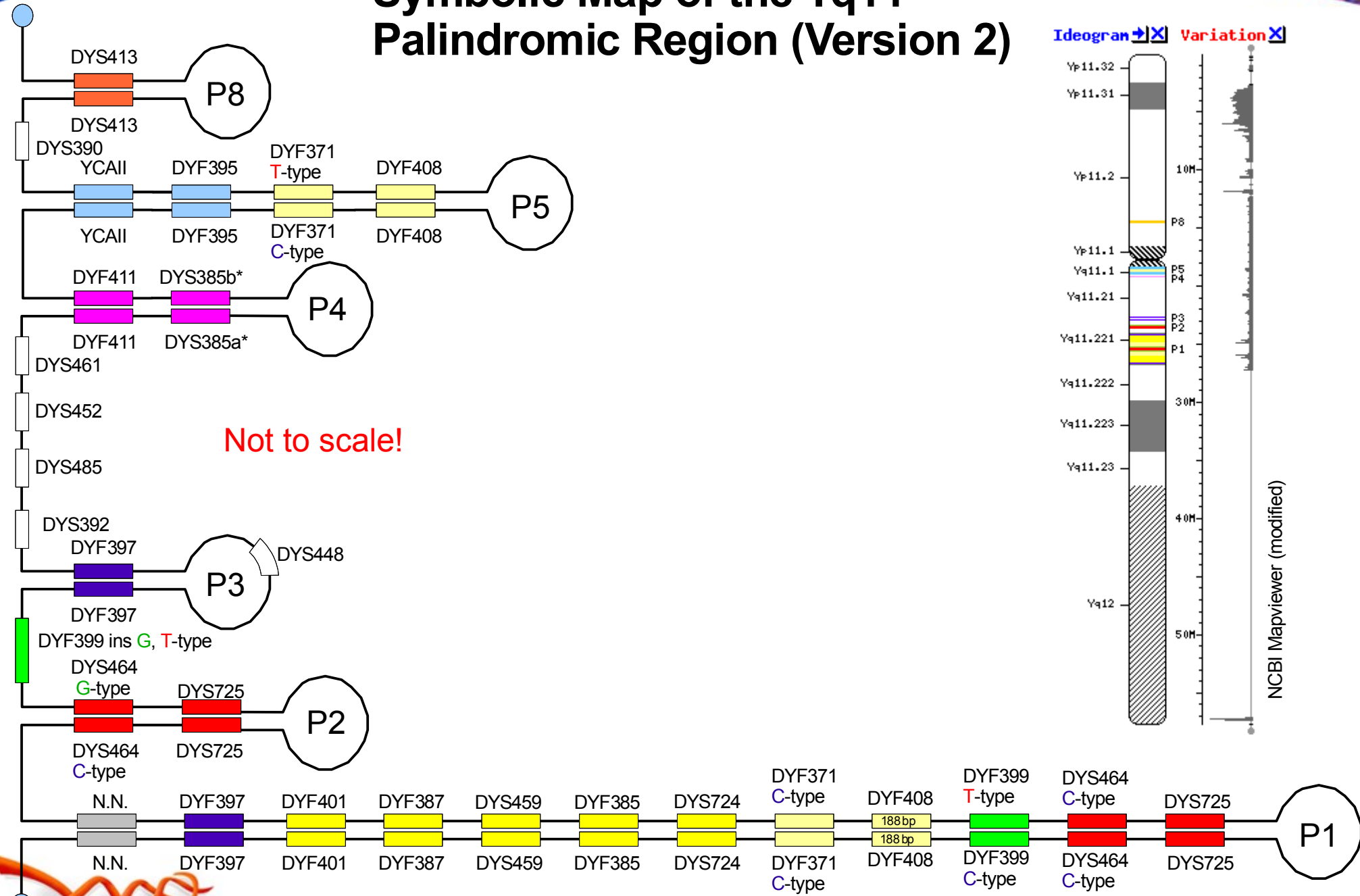
# Highest Density of Palindromes in the Human Genome

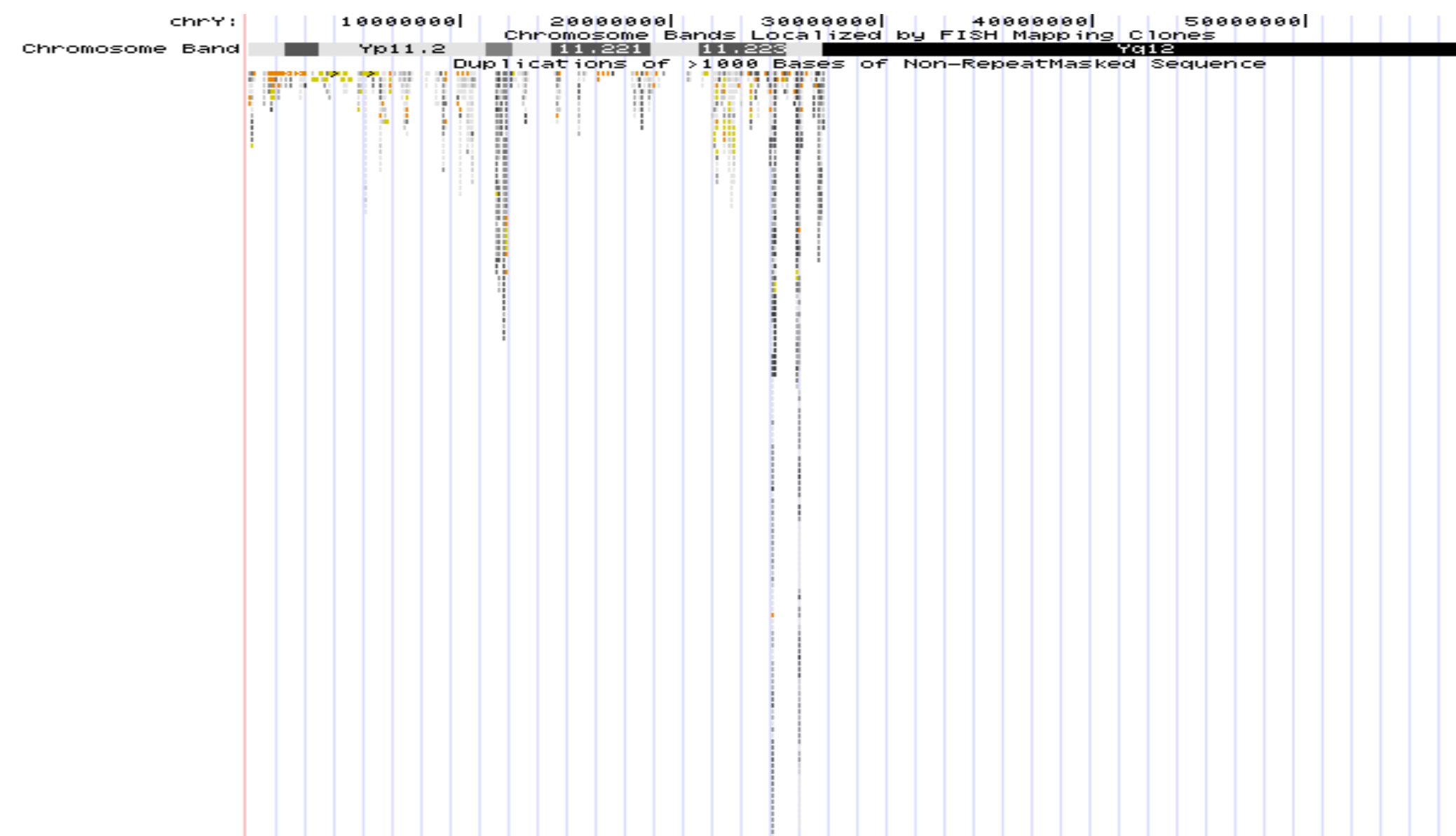
Segmental Duplications in the Human Genome

Chr	Length (bp)	Intrachromosomal		Interchromosomal		Total	
		Length	%	Length	%	Length	%
1	245,203,898	6,431,462	2.6	3,964,057	1.6	8,678,912	3.5
2	243,315,028	6,380,301	2.6	3,757,634	1.5	8,935,221	3.7
3	199,411,731	1,646,046	0.8	1,870,056	0.9	2,671,459	1.3
4	191,610,523	2,323,764	1.2	2,547,466	1.3	3,927,792	2.0
5	180,967,295	4,066,897	2.2	2,083,920	1.2	5,208,550	2.9
6	170,740,541	2,048,892	1.2	1,123,050	0.7	2,854,222	1.7
7	158,431,299	9,629,716	6.1	3,734,503	2.4	11,722,991	7.4
8	145,908,738	1,576,863	1.1	1,694,593	1.2	2,153,612	1.5
9	134,505,819	8,451,476	6.3	4,371,262	3.2	9,403,888	7.0
10	135,480,874	6,460,047	4.8	1,919,342	1.4	7,741,228	5.7
11	134,978,784	4,223,832	3.1	2,147,666	1.6	5,382,256	4.0
12	133,464,434	1,616,743	1.2	1,134,900	0.9	2,582,114	1.9
13	114,151,656	1,451,225	1.3	1,655,399	1.5	2,700,321	2.4
14	105,311,216	282,478	0.3	849,400	0.8	1,116,676	1.1
15	100,114,055	5,520,203	5.5	3,339,498	3.3	7,091,918	7.1
16	89,995,999	7,378,691	8.2	3,456,338	3.8	8,247,312	9.2
17	81,691,216	5,505,106	6.7	1,217,149	1.5	6,432,722	7.9
18	77,753,510	230,844	0.3	1,400,896	1.8	1,627,497	2.1
19	63,790,860	1,763,189	2.8	918,571	1.4	2,531,577	4.0
20	63,644,868	772,190	1.2	1,068,246	1.7	1,369,456	2.2
21	46,976,537	431,633	0.9	1,714,574	3.6	1,734,567	3.7
22	49,476,972	2,303,175	4.7	1,633,388	3.3	3,481,523	7.0
<b>Y</b>	152,634,166	3,579,325	<b>13.1</b>	4,550,908	3.0	8,047,172	5.3
	50,961,097	6,651,452		1,462,582	2.9	7,353,078	14.4
Total	3,070,521,116	90,725,550	3.0	53,615,398	1.7	122,996,064	4.0

Liqing Zhang et al. "Patterns of Segmental Duplication in the Human Genome", Mol. Biol. Evol. 22(1):135–141. 2005

# Symbolic Map of the Yq11 Palindromic Region (Version 2)





DYS459	6	7	8	9	10	11	12	13
6	4							
7	0	2						
8	3	9	404					
9	2	174	6467	7676				
10	2	120	2540	20908	554			
11	0	0	16	300	9	0		
12	0	0	0	4	0	0	0	
13	0	0	0	1	0	0	0	0

Scale:  0 1 10 100 1000 10000 100000

YCAII	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
11	0																
12	0	0															
13	0	0	0														
14	0	0	0	0													
15	0	0	0	0	0												
16	0	0	0	0	0	0											
17	0	0	0	0	0	0	11										
18	0	0	0	0	0	0	0	11									
19	3	1	0	0	7	1	19	15	826								
20	9	0	0	0	0	1	1	63	426	382							
21	72	2	1	0	0	3	21	90	4180	196	418						
22	11	0	0	0	0	0	3	45	2236	41	89	166					
23	0	0	0	0	1	1	33	251	12245	145	89	46	123				
24	0	0	0	0	0	0	1	2	748	8	7	7	40	0			
25	0	0	0	0	0	0	0	0	23	0	0	3	10	0	0		
26	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
27	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Scale: 0 1 10 100 1000 10000 100000

DYS385	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
6	0																					
7	0	0																				
8	0	0	0																			
9	0	1	0	6																		
10	0	0	0	0	46																	
11	0	0	0	9	87	1092																
12	0	0	1	14	36	817	720															
13	0	0	0	19	95	5181	1032	1063														
14	0	7	16	128	1143	27107	3455	4150	2976													
15	0	1	3	42	222	6037	1399	1843	2531	1428												
16	0	5	41	23	43	1128	511	1377	1161	1545	653											
17	0	2	5	4	44	210	250	1147	782	641	863	365										
18	0	0	1	3	67	107	172	600	417	404	990	892	216									
19	0	1	0	5	60	31	124	397	136	255	302	249	84	29								
20	0	0	0	1	43	13	83	131	88	91	104	40	25	9	1							
21	0	0	0	0	13	6	10	108	16	31	12	12	2	2	0	0						
22	0	0	0	0	3	0	2	39	8	5	5	8	1	0	1	0	0					
23	0	0	0	0	0	0	5	17	1	1	0	0	0	1	0	0	0	0				
24	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0			
25	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0		
26	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Scale: 0 1 10 100 1000 10000 100000





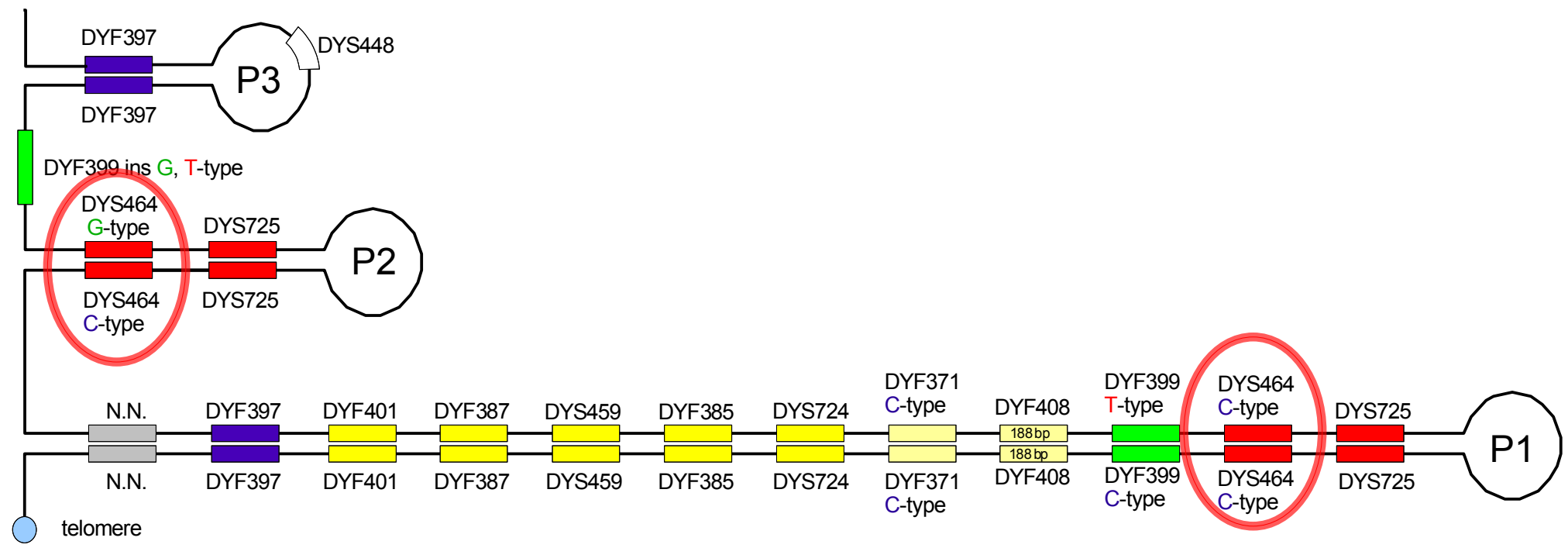
DYS724	9	26	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
9	0																			
26	0	0																		
28	0	0	0																	
29	0	0	0	1																
30	0	0	0	6	5															
31	0	0	0	5	5	11														
32	0	0	0	2	5	23	27													
33	0	0	0	4	29	55	53	61												
34	1	0	1	22	41	55	106	168	182											
35	0	0	0	8	38	80	95	229	467	404										
36	0	0	0	2	27	63	70	235	435	926	845									
37	0	1	1	2	14	90	75	173	547	1057	1846	1051								
38	0	0	0	1	17	46	45	199	433	939	1622	1963	935							
39	0	0	0	0	5	6	33	153	294	524	808	1151	957	290						
40	0	0	0	0	3	7	11	52	132	244	421	444	447	220	82					
41	0	0	0	0	1	4	12	16	58	114	138	132	134	83	40	12				
42	0	0	0	0	0	1	1	7	14	27	54	23	31	12	8	2	8			
43	0	0	0	0	0	1	1	0	1	7	8	7	4	4	1	1	0	1		
44	0	0	0	0	0	0	1	0	0	4	0	2	2	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Scale: 0 1 10 100 1000 10000

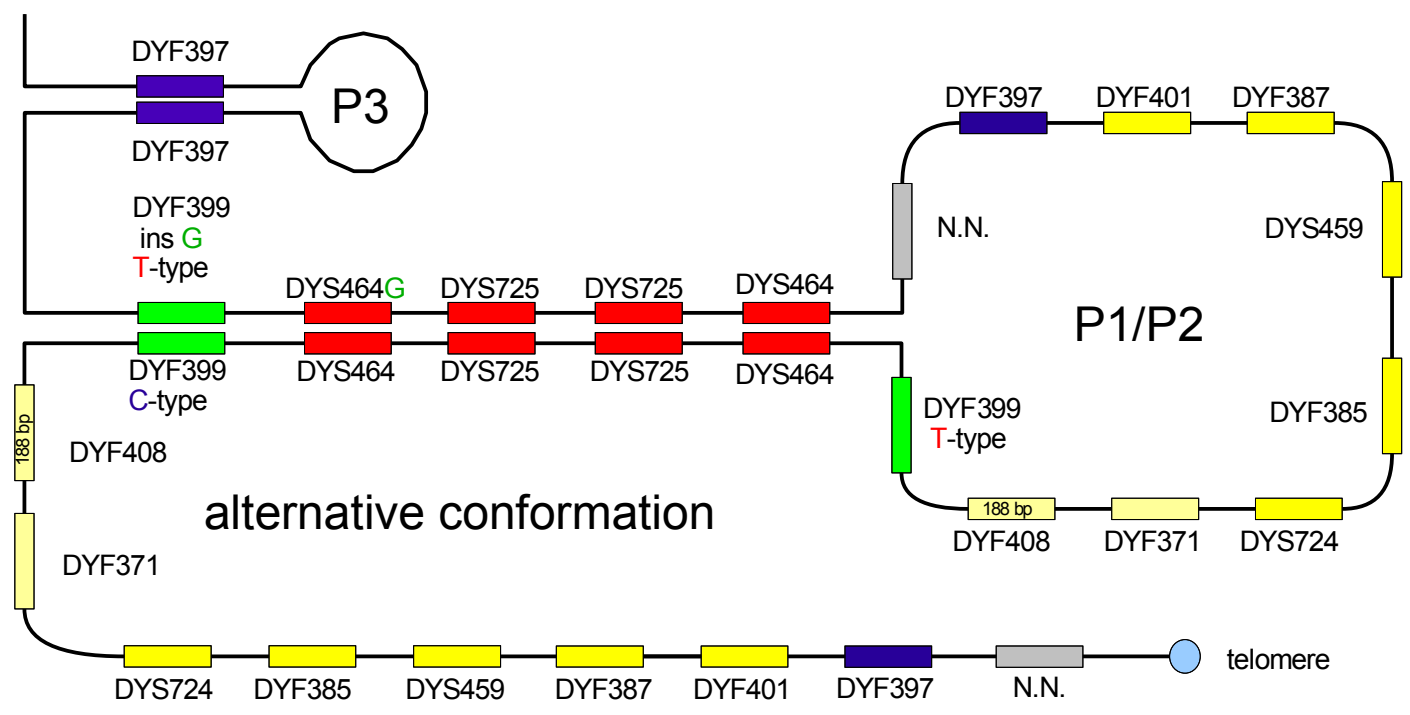




# DYS464 – A Complex Marker With (Usually) 4 Alleles



# DYS464 – A Complex Marker With (Usually) 4 Alleles





## DYS464 Extended PCR

If the primer ending with **G** yields a **peak**, the allele is classed as a **G**-type.

If the primer ending with **C** yields a **peak**, the allele is classed as a **C**-type.

DYS464-forward	<b>FAM</b> -ttacgagctttgggctatg (exactly like Redd paper)
DYS464-reverse- <b>G</b> -Type	<b>JOE</b> -cctgggtaacagagagactctttca <b>g</b> (ending with G)
DYS464-reverse- <b>C</b> -Type	<b>TMR</b> -cctgggtaacagagagactctttca <b>c</b> (ending with C)

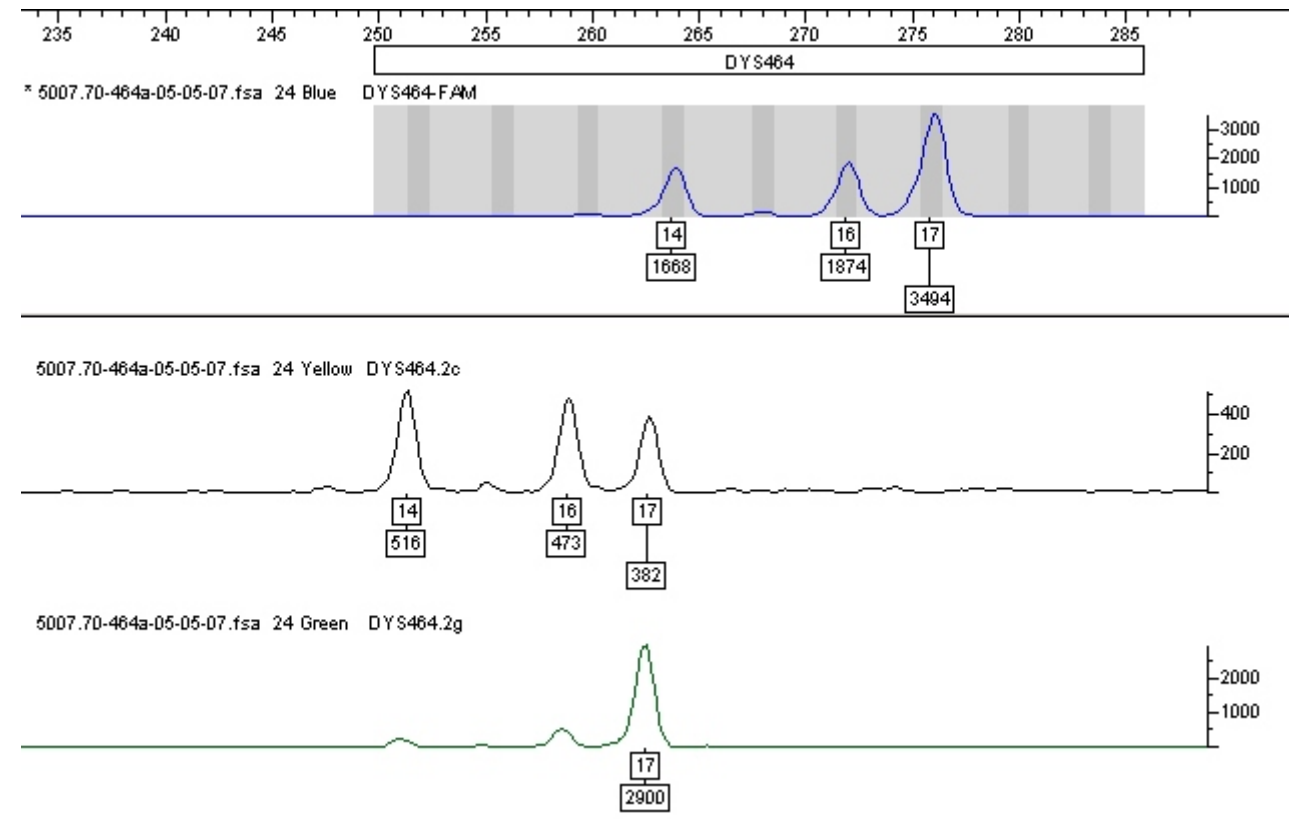
PCR parameters: 30 x (90°, 65°, 70°) and 60° end-elongation 1/2 h

Peaks with a **green fluorescence** (**JOE**) are called **G**-type.

Peaks with a **yellow fluorescence** (**TMR**) are called **C**-type.

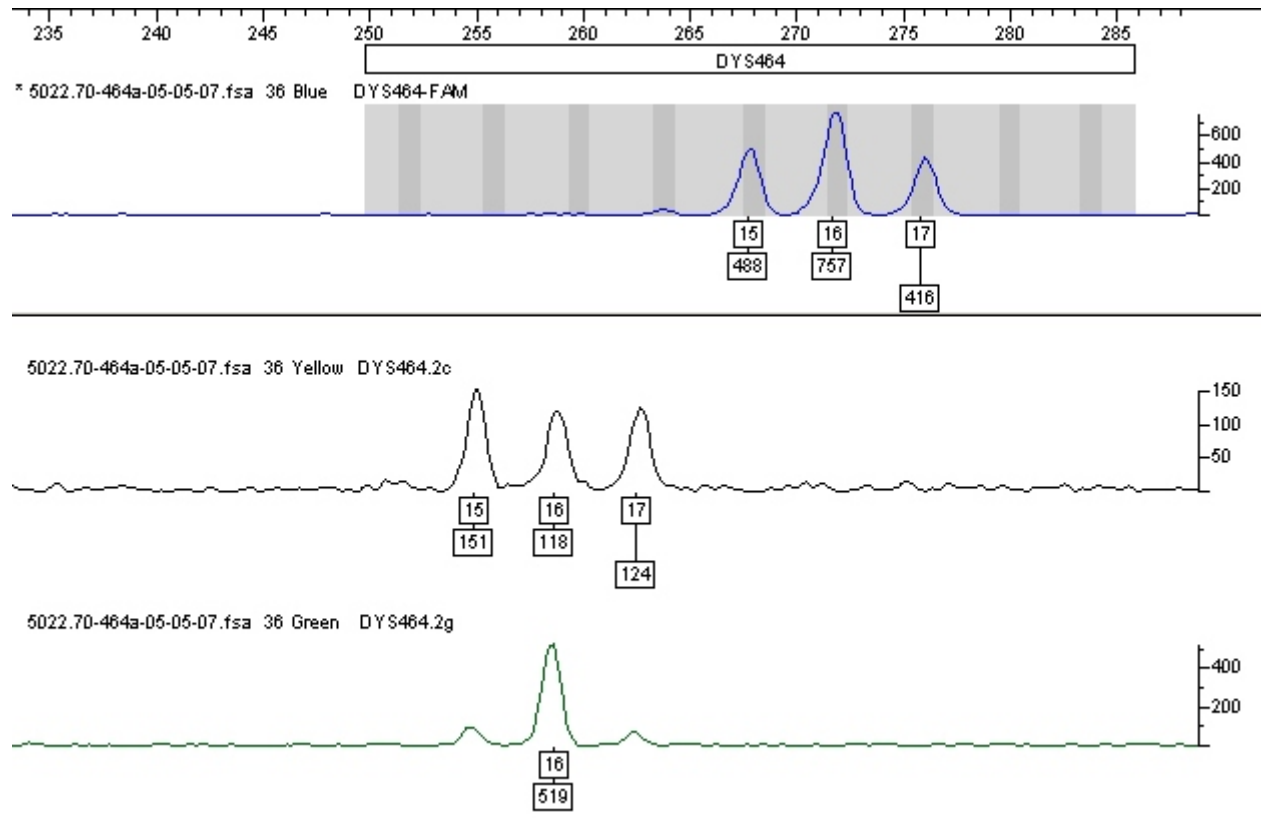
# DYS464 Extended PCR

## Electropherogram



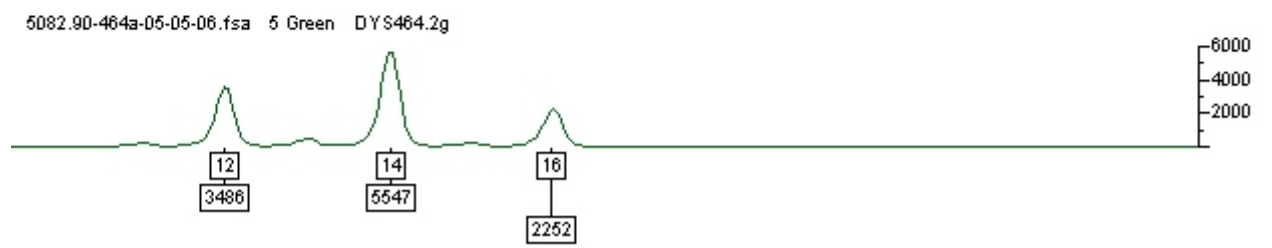
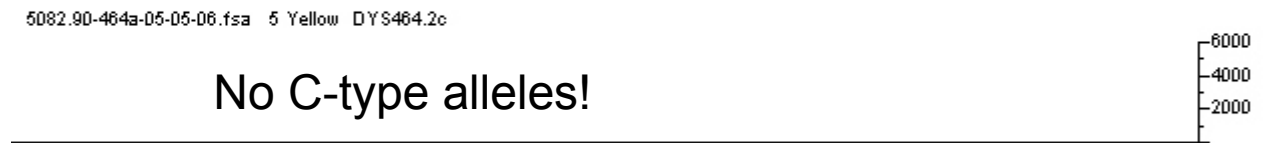
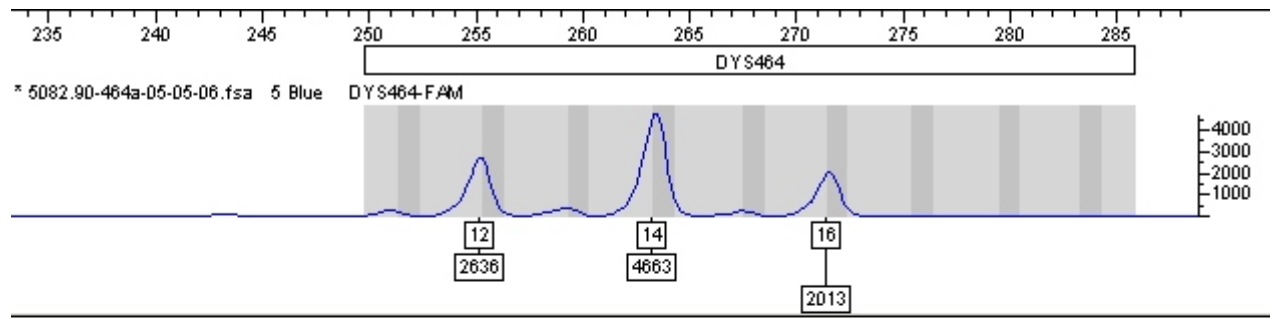
# DYS464 Extended PCR

## Electropherogram



# DYS464 Extended PCR

## Haplogroup I



# Typing of DYS464X

Y h group	DYS 464
A	11g-13g-13g-16g
E	14g-15.3g-17g-18g
E3b1	14g-15.3g-17g-18g
G	13g-14g-15g-15g
G2*	12g-12g-12g-13g
I	12g-14g-15g-16g
I1a	12g-14g-14g-16g
I1a3	12g-12g-14g-14g-15g-16g
I1b	11g-14g-14g-14g
I1b	11g-14g-14g-15g
I1b2a	11g-14g-14g-15g
I1b2a1	11g-11g-14g-15g
I1c	14g-15g-15g-16g
J2a1*	12g-13g-15g-16g-16g-16g
N	14g-14.3g
R1a1*	12g-15g-15g-16g
R1b	16c-16c-16g-16g
R1b	15c-16c
R1b	15c-15c-17c-17g
R1b	14c-16c-17c-17g
R1b	14c-15c-16g-17c
R1b	16c-16g
R1b	15c-15c-16c-16g
R1b	14c-15c-17c-17g
R1b	15c-16c-17g-17g
R1b	15c-16c
R1b	15c-15c-15c-15c
R1b	15c-15c-17c-17g
R1b	15c-17c-17c-18g
R1b	15c-15c-17c-18g
R1b	15c-15c-16g-17c
R1b	16c-16c-17c-17g
R1b	14c-15c-15c-15g
R1b	15c-15c-16c-18g
R1b	15c-15c-16c-17.1g
R1b	15c-15c-16g-17c
R1b	15c-15c-16g-17c
R1b	13c-15c-17c-17g
R1b	15c-15c-17g-17g
R1b	15c-16c-16c-18g
R1b	15c-15c-16c-17c

Other haplogroups have only **G**-type alleles



R1b has usually 3 **C**-type alleles and one **G**-type allele



Exceptions most likely products of recLOH



# What Is The Use Of The DYS464X Test?

Most R1b males that we have tested show 3 C-types and 1 G-type.  
All other haplogroups (including R1a) show 4 G-types.

- 1.) The DYS464X method can verify, if a person is really R1b
- 2.) In many cases we can determine, if a single large peak really represents two different alleles
- 3.) Some DYS464 patterns look similar when the conventional test is used, but really consist of completely different alleles. For example:

14c-15c-17c-17g is completely different from  
14c-15g-17c-17c,

but this wouldn't be seen with the conventional typing method.

- 4.) DYS464X is a method that helps us observe recombinational loss of heterocycosity (recLOH)



## DYS464 Examples From An Arbitrary Surname Project:

(conventional results)

surname A, 1: 14-14-16-17

surname A, 2: 14-14-16-17

surname A, 3: 14-14-16-17

surname B, 1: 14-14-16-18

surname C, 1: 14-14-15-17

If all the 5 persons are R1b then they can be typed with an extended DYS464 test.

We might get the following results:

surname A, 1: 14<sub>c</sub>-14<sub>c</sub>-16<sub>g</sub>-17<sub>c</sub>

surname A, 2: 14<sub>c</sub>-14<sub>c</sub>-16<sub>g</sub>-17<sub>c</sub>

surname A, 3: 14<sub>c</sub>-14<sub>c</sub>-16<sub>c</sub>-17<sub>g</sub>

surname B, 1: 14<sub>c</sub>-14<sub>c</sub>-16<sub>g</sub>-18<sub>c</sub>

surname C, 1: 14<sub>c</sub>-14<sub>c</sub>-15<sub>g</sub>-17<sub>c</sub>

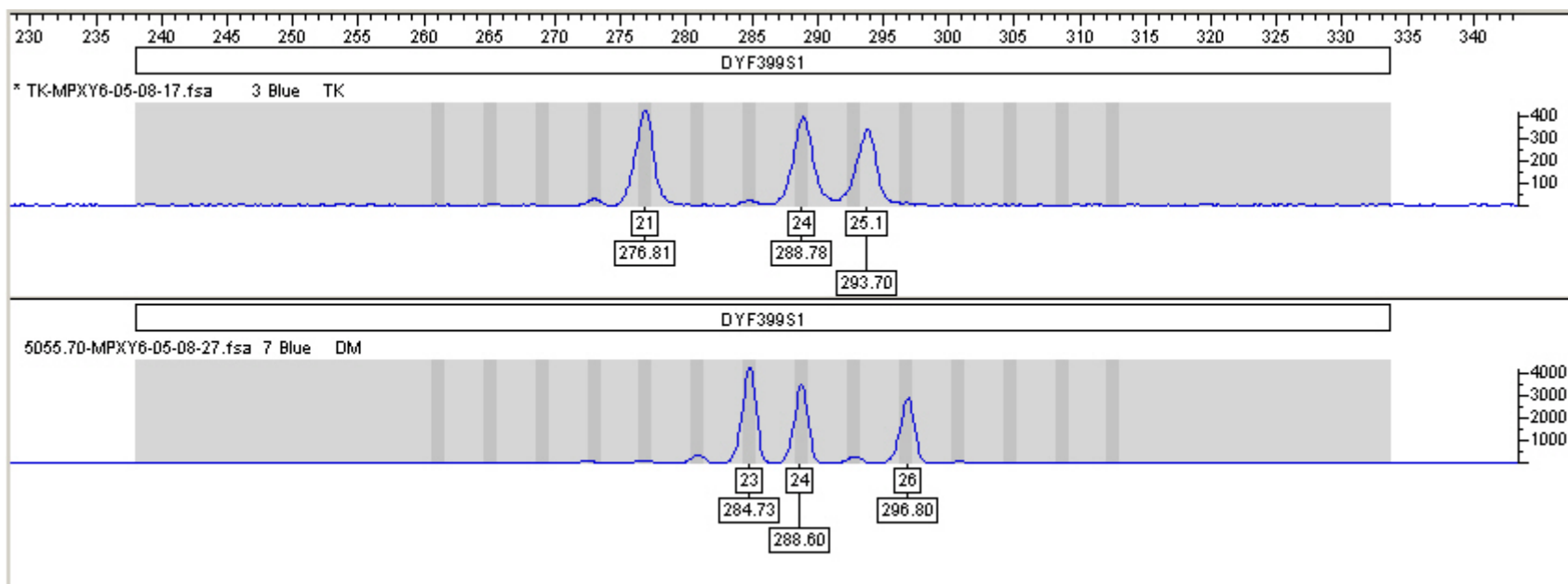
From that result we could say:

- A1 and A2 match completely
- A3 has at least a genetic distance of 2 from (A1 & A2)
- B1 has a genetic distance of 1 from (A1 & A2)
- C1 has a genetic distance of 1 from (A1 & A2)

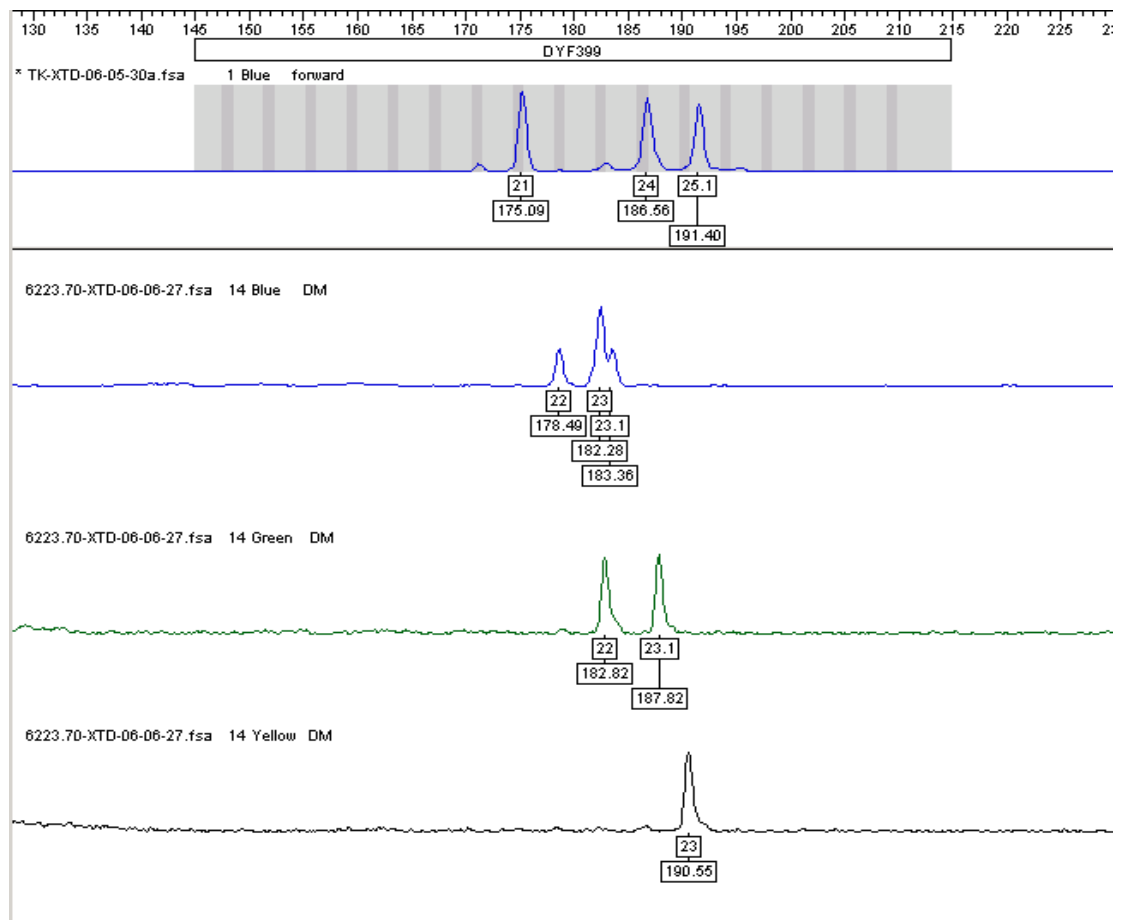


# DYF399

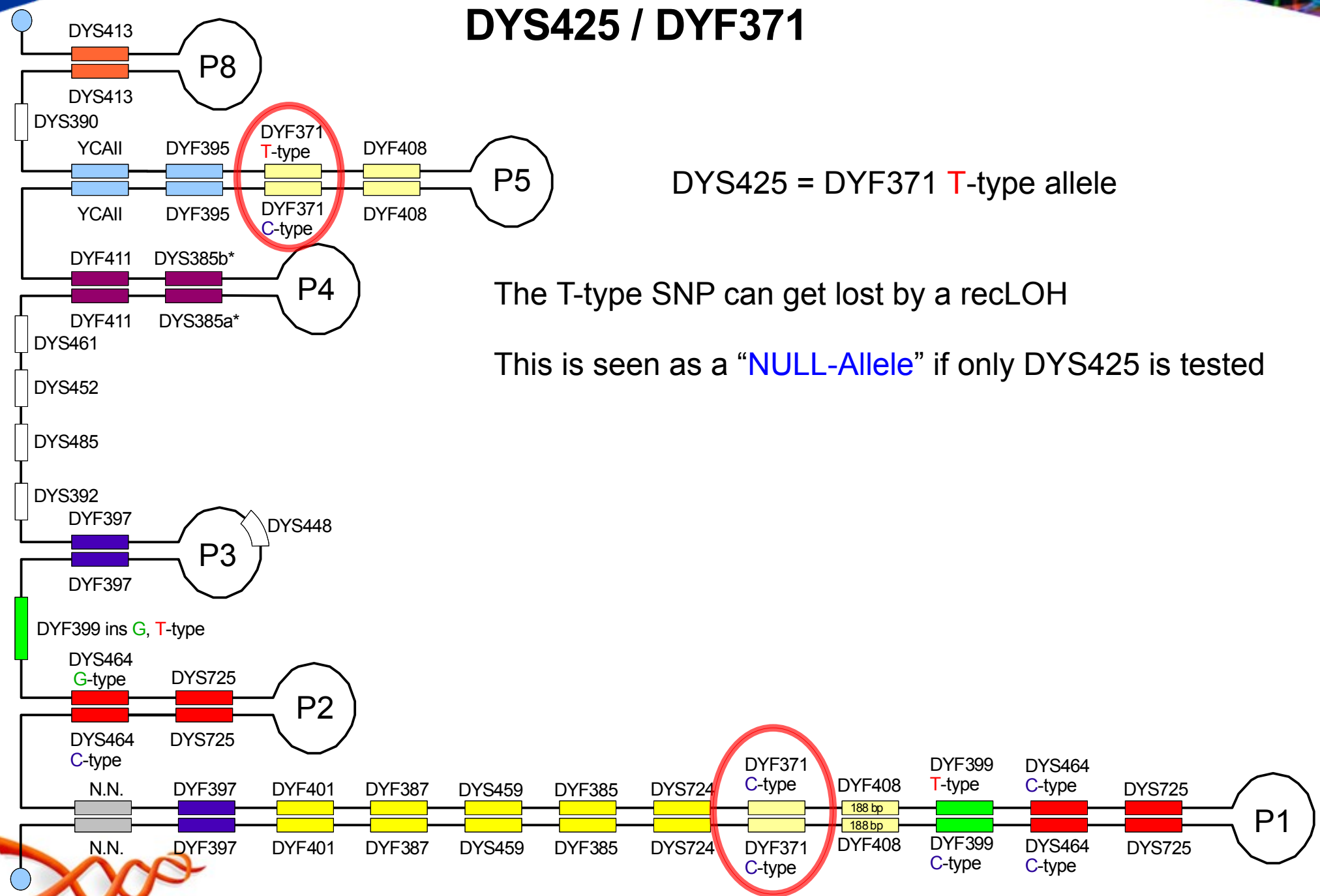
## A Fast Moving, Asymmetrical Palindromic Y-STR



# DYF399 - A Fast Moving, Asymmetrical Palindromic Y-STR



# DYS425 / DYF371



DYS425 = DYF371 T-type allele

The T-type SNP can get lost by a recLOH

This is seen as a "NULL-Allele" if only DYS425 is tested

# DYS425 / DYF371

The HUGO sequence has also a Null allele at DYS425

```
ttggagagaagaagagagaaaagggtcttagaatttagaaaaatgttgttgttggttggttggttggttggttggtt~taattttccattttaccctccagaattacttg
ttggagagaagaagagagaaaagggtcttagaatttagaaaaatgttggttggttggttggttggttggttggttggtt~taattttccattttaccctccagaattacttg
ttggagagaagaagagagaaaagggtcttagaatttagaaaaatgttggttggttggttggttggttggttggttggtt~taattttccattttaccctccagaattacttg
ttggagagaagaagagagaaaagggtcttagaatttagaaaaatgttggttggttggttggttggttggttggttggtt~taattttccattttaccctccagaattacttg
~TGGAGAGAAGAAGAGAGAAA~
-TGGAGAGAAGAAGAGAGAAAT
-TGGAGAGAAGAAGAGAGAAAC
~CCATTTTACCTCCAGAAATTACT
```

10c-10c-13c-14c

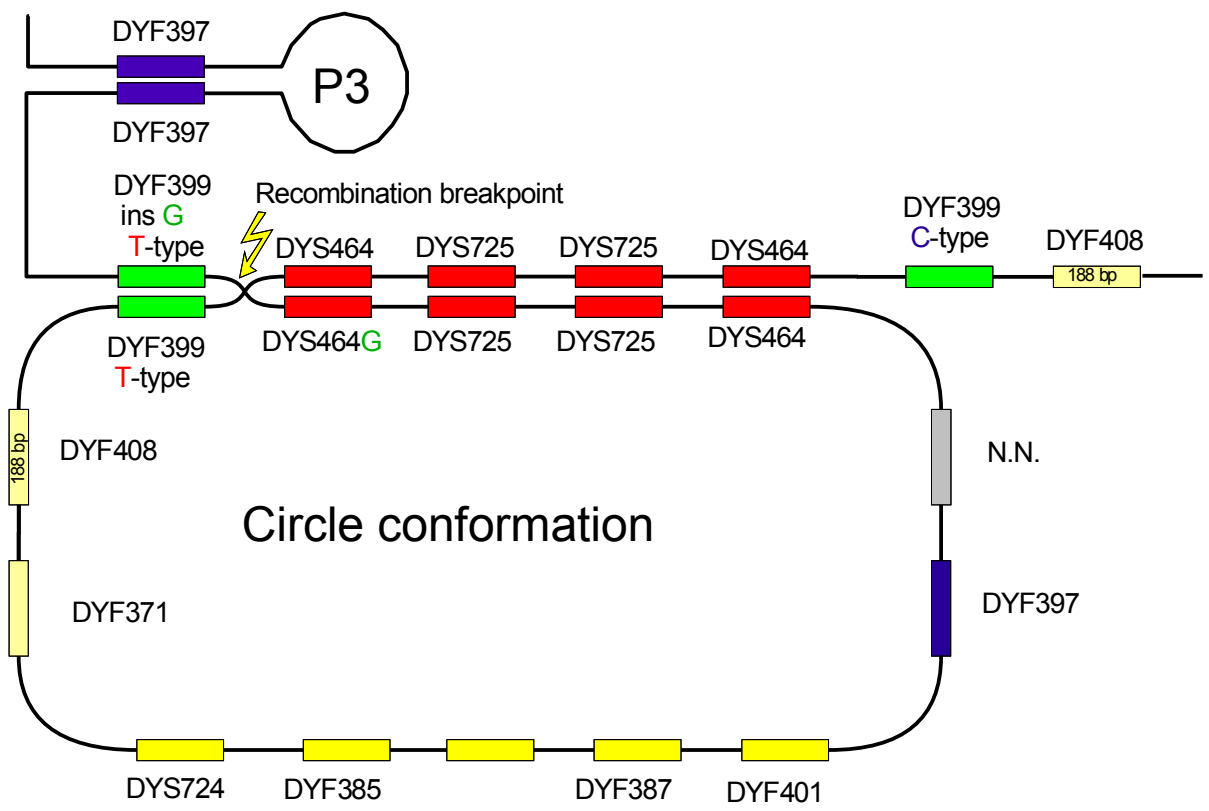
Normally in R1b (and most other haplogroups):

10c-12t-13c-14c



# How comes It To A Deletion?

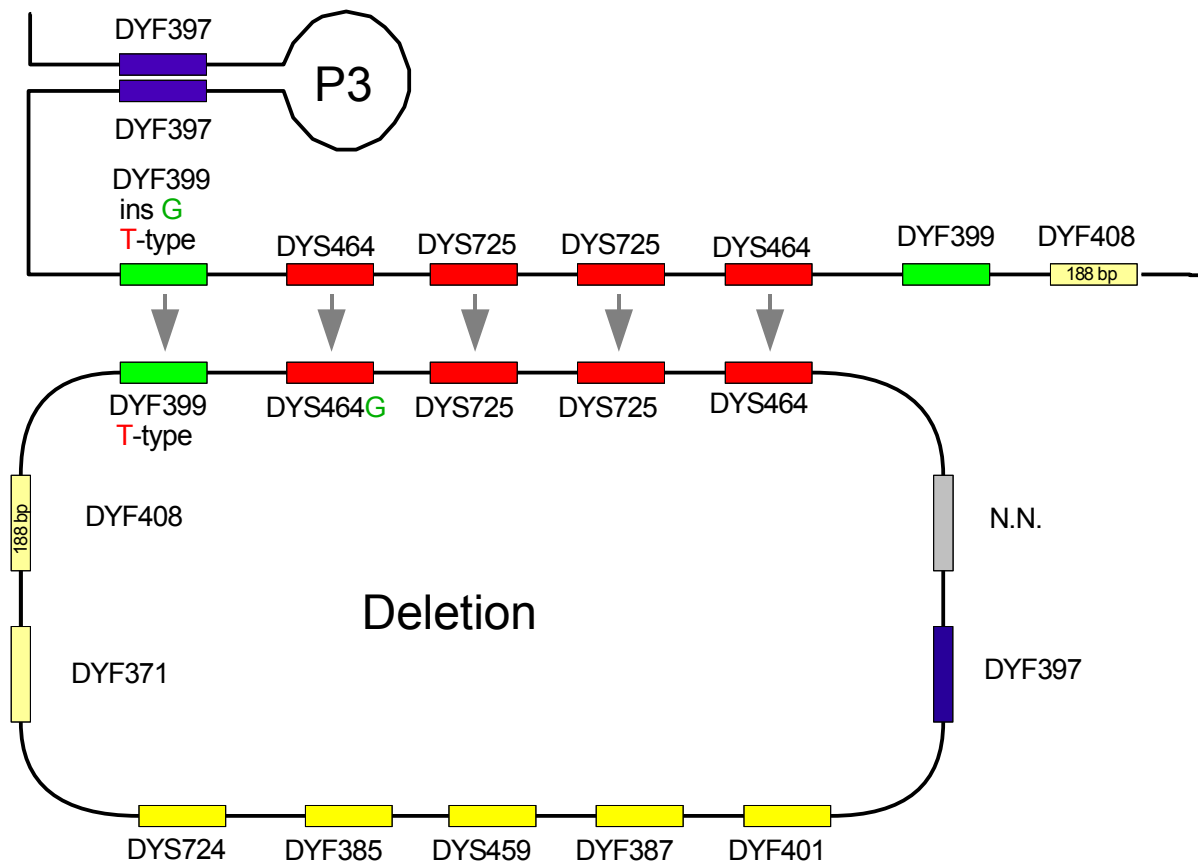
Symetry in the red/red (P1P2) region allows another irregular conformation:





# How comes It To A Deletion?

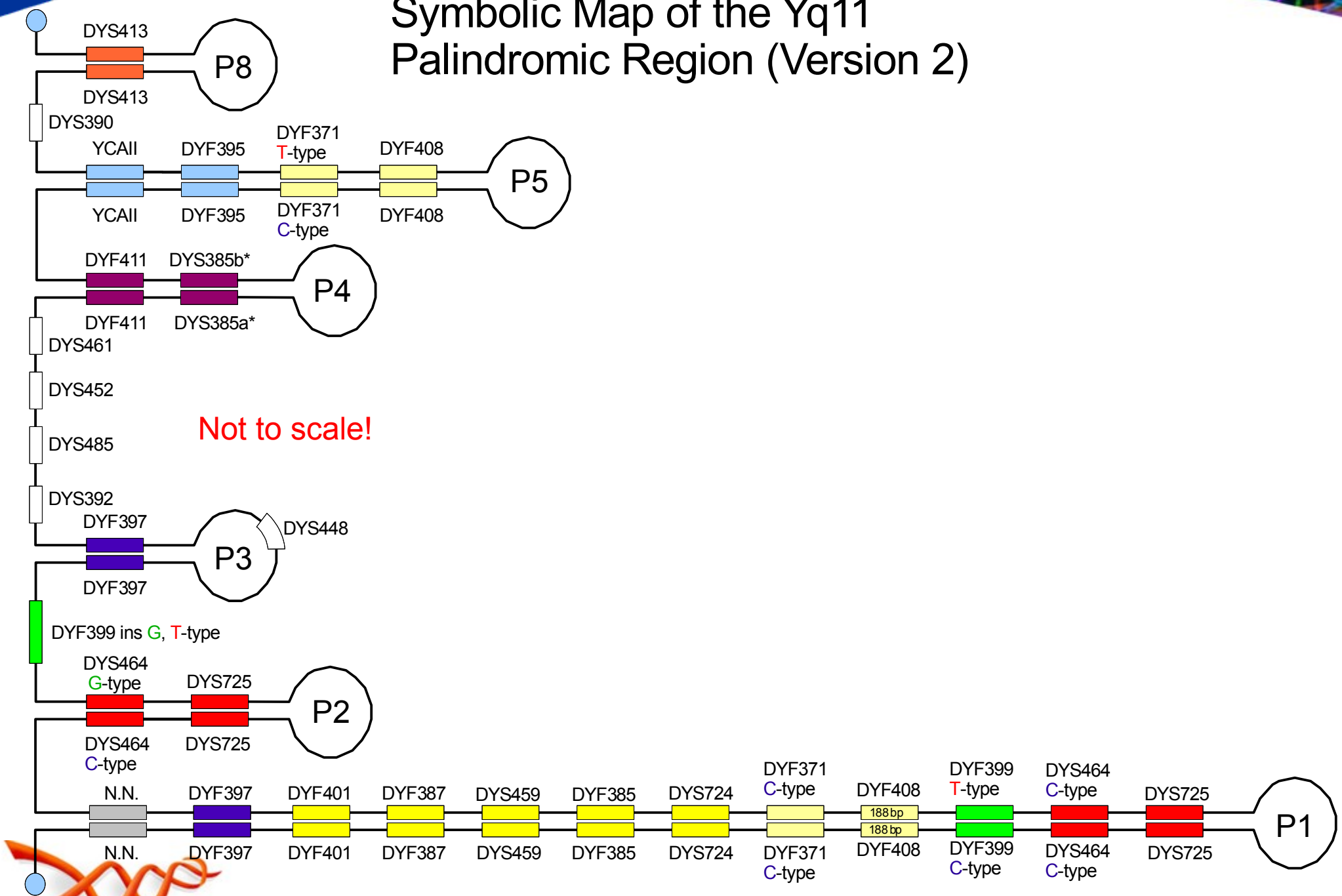
The circular DNA molecule can't replicate on its own and gets lost in the next cell cycle



# Famous People with a P1/P2 Deletion



# Symbolic Map of the Yq11 Palindromic Region (Version 2)



Not to scale!