International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:7, No:1, 2013

Evolutionary Distance in the Yeast Genome

Somayyeh Azizi, Saeed Kaboli, and Atsushi Yagi

***Abstract*—**Whole genome duplication (WGD) increased the number of yeast *Saccharomyces cerevisiae* chromosomes from 8 to 16. In spite of retention the number of chromosomes in the genome of this organism after WGD to date, chromosomal rearrangement events have caused an evolutionary distance between current genome and its ancestor. Studies under evolutionary-based approaches on eukaryotic genomes have shown that the rearrangement distance is an approximable problem. In the case of *S. cerevisiae*, we describe that rearrangement distance is accessible by using dedoubled adjacency graph drawn for 55 large paired chromosomal regions originated from WGD. Then, we provide a program extracted from a C program database to draw a dedoubled genome adjacency graph for *S. cerevisiae*. From a bioinformatical perspective, using the duplicated blocks of current genome in *S. cerevisiae*, we infer that genomic organization of eukaryotes has the potential to provide valuable detailed information about their ancestrygenome.

***Keywords*—**Whole-genome duplication, Evolution, Double-cut-and-join operation, Yeast.

I. Introduction

GENOME duplication is a fundamental process in the evolution of species. Within the yeast *Saccharomyces cerevisiae* (*S. cerevisiae*), whole-genome duplication (WGD), is thought to have occurred approximately 100 Ma, in which two identical copies of each chromosome with identical gene content and gene order created [1-5]. Since the WGD to date, organization of the genome of *S. cerevisiae* has been modifying under large chromosomal rearrangements which is critical for the evolution [6]. In other words, during the course of evolution, the genes in the genome can be shuffled around by genome rearrangements that move genes within a chromosome or among chromosomes. For multi-chromosomal genomes, the most common operations of rearrangement are intra-chromosomal translocations, inversions, non-reciprocal telomeric translocations, fusions, fissions [7] and block-interchanges [8] that all were initially introduced by Yancopoulos and colleagues in unique operation to sort two signed genomes which called *double-cut-and-join* (DCJ) [9]. It basically cuts a genome in two places and then joins the resulting four ends in a new way. Recently, Kahn and colleagues constructed an algorithm to compute duplication distance between two signed strings with segmental duplication. They also extended the computation for the genome with duplication-deletion and duplication-inversion-deletion in the human genome [10]. In [11], genome dedoubling problem was solved with introducing algorithm in the DCJ and reversal models for reconstruction of non-duplicated ancestor of *Drosophila yakuba*. In [12], Chen X and colleagues found how to sort unsigned genome composed of linear and circular chromosome by double-cut-and-join operations. In this paper we calculate the minimum number of rearrangement needed to transform present-day genome of *S. cerevisiae* to its ancestor (DCJ distance) by using theorem 1 mentioned in [11]. To approximate the evolutionary distance between two genomes, we first use updated map of duplicated regions in the yeast genome [2] to draw dedoubled genome adjacency graph. It is well known that the genome of *S. cerevisiae* contains 55 different pairs of syntenic chromosomal regions originated from WGD, consisting of duplicated gene pairs arranged in the same order on two different chromosomes or rarely on a single chromosome [1, 13]. Then, we present a program which is applicable to draw dedoubled genome adjacency graph of *S. cerevisiae* and also any uni- or multi- chromosomal signed genome.

Somayyeh Azizi is now with Department of Applied Physics, Graduate School of Engineering, Osaka University, Osaka, Japan (e-mail: somayyeh_azizi@ap.eng.osaka-u.ac.jp).

Saeed Kaboli is now with Department of Biotechnology, Graduate School of Engineering, Osaka University, Osaka, Japan. Department of Bioscience, School of Science, Zanjan University, Zanjan, Iran (corresponding author e-mail: kaboli.saeed@bio.eng.osaka-u.ac.jp).

Atsushi Yagi is now with Department of Applied Physics, Graduate School of Engineering, Osaka University, Osaka, Japan (e-mail: yagi@ap.eng.osaka-u.ac.jp).

II. Material and Methods

A. Whole Duplicated Genome

Each linear chromosome of the genome is composed of duplicated segments and their paralogs which are represented by signed genomic markers. Two copies of a same marker in a genome are called paralogs. Markers are indicated by sign in which the sign shows orientations of markers in chromosomes. Two ends of each linear chromosome indicate telomeres which are represented by the unsigned marker ○. If a marker x is present twice, one of the paralogs is represented by $\underline{x}$. A duplicated genome is a genome in which subset of markers are duplicated. For example, (○ 2 $\underline{1}$ -5 6 ○) (○ $\underline{2}$ -4 ○) (○ -1 $\underline{6}$ 3 $\underline{5}$ ○) is a multi-chromosomal genome composed of four duplicated markers and their paralogs. Whole duplicated genome is a duplicated genome in which all segments are duplicated e.g., (○ 4 -5 $\underline{-5}$ ○) (○ 2 $\underline{4\text{-}2}$ ○). Whole duplicated genome in ancestor is a genome such that marker *x* and its paralog are in form of pairs (*x* $\underline{x}$) or ($\underline{x}$ *x*). For example, (○ 1$\underline{1}$ 2 $\underline{2}$ 3 $\underline{34}$ 4 5 $\underline{5}$ ○).

B. Whole Duplicated Genome in the S. cerevisiae

A *perfectly doubled* genome of *S. cerevisiae* included in 16 linear chromosomes rearranged over evolutionary time to construct current genome with cluttered gene order. Present-day rearranged genome of *S. cerevisiae* carrying 55 duplicated pair regions which are surrounded by single-copy genes within

chromosomes. An overview of location of the 55 duplicated regions on the 16 chromosomes is depicted in Fig. 1.

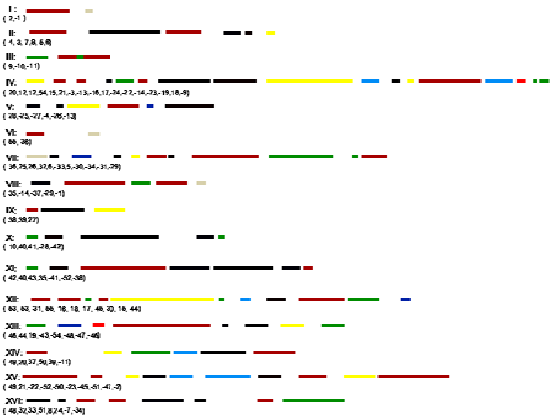

Fig. 1 Locations of the 55 duplicated regions on the 16 chromosomes of the yeast genome

A dedoubled genome of *S. cerevisiae* isa duplicated genome *G(Sc)* such that for any duplicated marker *x* in *G(Sc)*, either (*x* $\underline{x}$), or ($\underline{x}$ *x*) is an adjacency of *G(Sc).*

C. Rearrangement by DCJ Operation

In this paper rearrangement is presented as DCJ operation that acts on two adjacencies *ab*and *cd* of a genome, transforms them either to the two adjacencies *ac* and *bd* or two adjacencies *ad* and *bc*. For example, the following DCJ cuts adjacencies (2 6) and (-5 2) to produce adjacencies (2 2) and (-5 6). Here, breakpoints are indicated by symbol ▲,and new adjacencies are presented by dots. (○ 2▲ 6 -3 7 ○)(○ 4-5▲ 2 -3 5 ○) → (○ 4-5 . 6 -3 7 2 . 2 -35 ○)

III. Results

We first study the *S. cerevisiae* genome dedoubling problem under the DCJ model. Then, we present a program which is applicable to draw dedoubled adjacency graph of genomes as well as the genome of *S. cerevisiae*.

A. Dedoubleing of S. cerevisiae Genome by DCJ Operation

In order to express the DCJ dedoubled distance, $d_{dcj}(Sc)$, between two whole duplicated genomes, we use a graph called dedoubled adjacency graph of *S. cerevisiae*.

B. Dedoubled Adjacency Graph in S. cerevisiae

The dedoubled adjacency graph of *S. cerevisiae* denoted by *A*(*Sc*), is the graph consists of two edges for each marker of *G(Sc)*, one edge between the vertices (*x u*) and (*v* $\underline{x}$), another between the vertices (*y x*) and ($\underline{x}$ *z*). For example, marker -13 into the *G(Sc)* contains two edges between the vertices (-13 -16), (-26 -13) and (3 -13), (-13○) respectively (see Fig. 2).

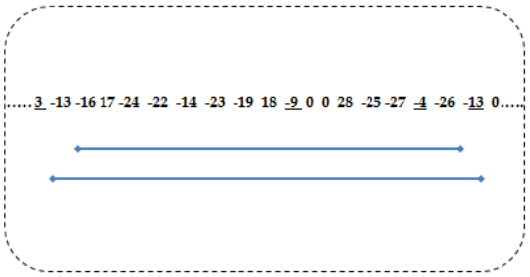

Fig. 2 Vertices and edges for marker -13 in *A (Sc)*

Vertices are the adjacencies of *G(Sc)*. Note that all vertices in *A(Sc)* have degree one or two. Thus, the connected components of *A(Sc)* are only paths and cycles. These paths and cycles are called elements of *A(Sc)*. If an element contains twice the couple (*x*, $\underline{x}$) so is called a duplicated element of *A(Sc)*, otherwise is called non-duplicated, namely if an element of *A(Sc)* contains no couple (*x*, $\underline{x}$) twice, then it is non-duplicate elements of this graph. If two edges of each marker belong to two different elements of *A(Sc)*, so both elements contain couple (*x*, $\underline{x}$) and then these two element are intersect, Otherwise are called independent.

Post-WGD information of *S. cerevisiae* (*http://acer.gen.tcd.ie/~khwolfe/yeast)* helped us to draw dedoubled adjacency graph of the present-day genome of budding yeast with 55 pairs of chromosomal markers as depicted in Fig. 3.

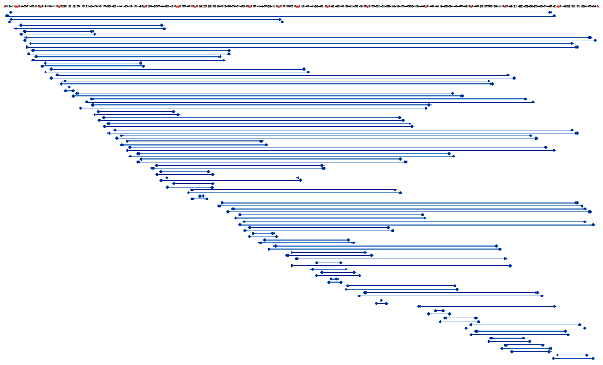

Fig. 3 The dedoubled adjacency graph of *Saccharomyces cerevisiae* genome

Based on the drawn graph, 16 elements were resulted in which 13 elements are paths while remaining 3 elements are cycles. All constructed cycles, i.e. C1;{(12, $\underline{12}$)}, C2;{(53, $\underline{53}$)} and C3;{($\underline{14}$ -37) ($\underline{37}$ 50) (-$\underline{50}$ -$\underline{23}$)} = (-37, $\underline{37}$) (50, $\underline{50}$)} are non-duplicated cycles. None of these non-duplicated cycles have common couple (*x*, $\underline{x}$) together. Thus, they are non-duplicated pairwise independent cycles.

According to theorem 1 of [11], for *S. cerevisiaed*$_{dcj}(Sc)$ would be equal with $n - C_i$. In this equation, let n be the number of couple of paralogous marker in *G(Sc)* and let $C_i$be the maximum size of subset of non-duplicated pairwise independent cycles in *A(Sc)*. Thus, according to Fig. 3

```
#include <conio.h>
#include<iostream.h>
struct numbers
{
    int n;
intloc;
intloc_b;
};
int main()
{
clrscr();
intnum[] = {0,  2, -1 ,0 ,0 , 4, -3, -7,  8, -5,  6 ,0 ,0 , 9, -10, -11 , 0 , 0 , 20, 12, 12, 54, 15, 21, -3, -13, -16, 17, -24, -22, -14, -23, -19, 18, -9 ,0 , 0 ,28, -25, -27, -4, -26, -13 , 0, 0 , 55, -36 ,0 , 0 , 36, 25, 26, 32,  6, -33, 5, -30, -34, -31, -29, 0 ,0 , 35, -14, -37, -29, -1, 0, 0, 38, 39,  27,0 ,0 ,10, 40, 41, -28, -42 ,0 ,0 , 42, 40, 43, 35, -41, -52, -38, 0 ,0 , 53, -53, -31, -55, -16, -18, -17, -45, -30, -15, -44, 0 , 0, 46, 44, 19, -43, -54, -48, -47, -46, 0,  0,  49,  20,   37,  50,   39,  -11,  0,  0,  49, 21, -22, -52, -50, -23, -45, -51,  -47,  -2,  0,  0, 48, 32, 33, 51, 8, 24, -7, -34 , 0};
int size=sizeof(num)/sizeof(int),i,j,k,n=0,num1[100],nu[100],location,location_b,k1;
numbers x[150];
 /*for(i=0;i<size;i++)
 {
    if(num[i]!=0)
  {
       x[n].n= num[i]
       x[n].loc=i*2;
n++;
  }
 } */
  k=0;
for(i=0;i<size;i++)
    for(j=0;j<size;j++)
        if(num[i]+num[j]==0 &&num[i]>0)
    {
    //cout<<num1[i]<<"    "<<num1[j];
nu[k]=num[i];
x[k].n=num[i];
x[k].loc=i*2;
x[k].loc_b=j*2;
k++;
       // cout<<"\n";
break;
    }
int ii=0,temp;
for(i=0;i<size;i++)
 {
    if(num[i]<-9)
        cout<<num[i];
if((num[i]>-9 &&num[i]<0)  || (num[i]>9))
  {
    cout<<num[i]<<" ";
  }
if(num[i]>=0 &&num[i]<10)
cout<<num[i]<<"  ";
 }
/*for(i=0;i<k;i++,ii+=2)
 {
    for(j=0;j<k;j++)
  {
    if(nu[i]==x[j].n)
    {
    //cout<<"\n";
location = x[j].loc;
location_b = x[j].loc_b;
if(location>location_b)
       {
temp=x[j].loc;
x[j].loc=location_b;
location_b=temp;
temp=location;
location=x[j].loc_b;
x[j].loc_b=temp;
       }
if(location>80)
    location %= 80;
location++;
gotoxy(location,i+8);
    cout<<"*";
while(location+1<location_b-1)
       {
    gotoxy(location+1,i+8);
        cout<<"_";
location++;
       }
gotoxy(location_b-1,i+8);
    cout<<"*";
location= x[j].loc;
location_b= x[j].loc_b;
gotoxy(location-1,i+9);
cout<<"*";
while(location<=location_b)
       {
    gotoxy(location,i+9);
        cout<<"_";
location++;
       }
gotoxy(location,i+9);
cout<<"*";
break;
    }
  }
}*/
for(i=0;i<k;i++)
 {
    for(j=0;j<k;j++)
  {
    if(nu[i]==x[j].n)
    {
location = x[j].loc;
location_b = x[j].loc_b;
if(location>location_b)
       {
temp=x[j].loc;
x[j].loc=location_b;
location_b=temp;
temp=location;
location=x[j].loc_b;
x[j].loc_b=temp;
       }
cout<<"\n";
       k1=0;
while(k1<location%80)
       {
        cout<<" ";
k1++;
       }
cout<<"*";
while(location<location_b-2)
       {
        cout<<"___";
location++;
       }
cout<<"*"<<endl;
location = x[j].loc;
       k1=0;
while(k1<(location%80)-2)
       {
        cout<<" ";
k1++;
       }
cout<<"*";
while(location<=location_b)
       {
        cout<<"___";
location++;
       }
cout<<"*";
break;
    }
  }
 }
getch();
return 0;
}
```

Fig. 4 Program for drawing dedoubledadjacency graph of *Saccharomyces cerevisiae* genome

indicating adjacency graph, n = 55 and C_i = 3. Eventually, the DCJ dedoubling distance of *S. cerevisiae* would be $d_{dcj}(Sc)$ = 55 - 3 = 52.

Program to draw dedoubled adjacency graph of S. cerevisiae: In this section, we introduce a program (Fig. 4) extracted from a C program database to draw dedoubled adjacency graph of signed genome of *S. cerevisiae*. It draws this graph by constructing one edge between the vertices (x u) (v $\underline{x}$) and one edge between the vertices (y x) and ($\underline{x}$ z) for any marker x. The program is also applicable to draw the graph for any signed genome.

IV. Conclusion

From a bioinformatical perspective, using the duplicated blocks in the current genome of *S. cerevisiae*, we have inferred that how to generalize the dedoubled adjacency graph for whole genome of this organism which is valuable for deep understanding of rearrangement events since WGD to date. In this work, we have shown that the number of non-duplicated pairwise independent cycle of current genome negatively effects on its distance with ancestor genome.

References

[1] Wolfe KH, Shields DC, "Molecular evidence for an ancient duplication of the entire yeast genome", Nature 387, 708±713, 1997.
[2] Seoighe C, Wolfe KH, "Updated map of duplicated regions in the yeast genome", Gene 238, 253–261, 1999.

[3] Kellis M, Birren BW, Lander ES, "Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae", Nature 428: 617-624, 2004.
[4] Vitkup D, Kharchenko P, Wagner A, "Influence of metabolic network structure and function on enzyme evolution", Genome Biol 7: R39,2006.
[5] Conant GC, Wolfe KH, "Turning a hobby into a job: how duplicated genes find new functions", Nat Rev Genet 9: 938–950, 2008.
[6] Gordon JL, Byrne KP, Wolfe KH, "Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern Saccharomyces cerevisiae genome", Plos Genet 5(5) e1000485, 2009.
[7] Gordon JL, Byrne KP, Wolfe KH, "Mechanisms of chromosome number evolution in yeast", Plos Genet 7(7) e1002190, 2011.
[8] Christie, D.A., "Sorting by block-interchanges", Inf. Process. Lett. 60, 165–169, 1996.
[9] Yacopoulos S, Attie O, Friedberg R, "Efficient sorting of genomic permutations by translocation, inversion and block interchange", Bioinformatics 21(16): 3340-3346, 2005.
[10] Kahn CL, Mozes S, Raphael B J, "Efficient algorithms for analyzing segmental duplications with deletions and inversion in genomes", Algorithms for Molecular Biology 5: 11, 2010.
[11] Thomas A, Varré JS, Ouangraoua A, "Genome dedoubling by DCJ and reversal", BMC bioinformatics 12(Suppl 9):S20, 2011.
[12] Chen X, Sun R, Yu J, "Approximating the double-cut-and-join distance between unsigned genomes", BMC Bioinformatics 12 (Suppl 9): S17, 2011.
[13] Mewes HW, Albermann K, BaÉhr M, Frishman D, Gleissner A et al., "Overview of the yeast genome", Nature 387, Suppl., 7±65,1997.

SomayyehAzizi was born in Zanjan, Iran. She received the B. D. in the Department of Computer Engineering from Zanjan Azad University in 2010. She is currently in the five years doctor's course in the Department of Applied Physics, Osaka University. She is interested in Bioinformatics.

Saeed Kaboli was born in Zanjan, Iran on August 23, 1978. He received the M.S. degree in the Department of Medical Mycology and Parasitology from Mazandaran University of Medical Sciences, Sari, Iran in 2008. He is currently in the doctor's course in the Department of Biotechnology, Graduate School of Engineering, Osaka University, Osaka, Japan. He is interested in Fungal Genetics and Evolutionary Genomics of Industrially and Medically Important Fungi.

Atsushi Yagi was born in Niigata, Japan in 1951. He received the Ph.D. degree in Division of Natural Science from Osaka University in 1982. He is currently a full Professor of Department of Applied Physics, Graduate School of Engineering, Osaka University, Osaka, Japan. His research interests include Evolution Equations, Self-Organization Model and its application in Physics, Chemistry and Biology.