# turnitin 🕖

## **Digital Receipt**

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Hongki Julie
Assignment title:	Periksa similarity
Submission title:	The distribution of genome shared identical by descent for
File name:	ir_of_full_sibs_by_means_of_the_continuous_time_Markov_ch
File size:	385.18K
Page count:	6
Word count:	2,859
Character count:	12,765
Submission date:	07-Apr-2022 11:24PM (UTC+0700)
Submission ID:	1804416985



Copyright 2022 Turnitin. All rights reserved.

# The distribution of genome shared identical by descent for a pair of full sibs by means of the continuous time Markov chain

Submission date: 07-Apr-2022 11:24PM (UTC+0700) Submission ID: 1804416985 File name: ir\_of\_full\_sibs\_by\_means\_of\_the\_continuous\_time\_Markov\_chain.pdf (385.18K) Word count: 2859 Character count: 12765



## The distribution of genome shared identical by descent for a pair of full sibs by means of the continuous time Markov chain

Hongki Julie, Udjianna S. Pasaribu, and Adi Pancoro

Citation: AIP Conference Proceedings **1692**, 020007 (2015); doi: 10.1063/1.4936435 View online: http://dx.doi.org/10.1063/1.4936435 View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1692?ver=pdfcov Published by the AIP Publishing

Articles you may be interested in

Efficient maximum likelihood parameterization of continuous-time Markov processes J. Chem. Phys. **143**, 034109 (2015); 10.1063/1.4926516

The evolution of the market share of competing brands modeled with Markov chains AIP Conf. Proc. **1648**, 670006 (2015); 10.1063/1.4912901

Calculating Mean First Passage Times from Markov Models of Proteins AIP Conf. Proc. **940**, 150 (2007); 10.1063/1.2793397

Graph transformation method for calculating waiting times in Markov chains J. Chem. Phys. **124**, 234110 (2006); 10.1063/1.2198806

An adjustable aperiodic model class of genomic interactions using continuous time Boolean networks (Boolean delay equations) Chaos **13**, 1167 (2003); 10.1063/1.1608671

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 120.164.40.190 On: Thu, 24 Dec 2015 03:53:43

### The Distribution of Genome Shared Identical By Descent For A Pair Of Full Sibs By Means Of The Continuous Time Markov Chain

Hongki Julie<sup>1</sup>, Udjianna S. Pasaribu<sup>2</sup>, and Adi Pancoro<sup>3</sup>

<sup>1</sup>Mathematic Education Department, Sanata Dharma University. <sup>2</sup>Mathematics Department, Institute Technology of Bandung. <sup>3</sup>Biology Department, Institute Technology of Bandung. \*Email: hongkijulie@yahoo.co.id

Abstract. This paper will allow Markov Chain's application in genome shared identical by descent by two individual at full sibs model. The full sibs model was a continuous time Markov Chain with three state. In the full sibs model, we look for the cumulative distribution function of the number of sub segment which have 2 IBD haplotypes from a segment of the chromosome which the length is *t* Morgan and the cumulative distribution function of the number of sub segment which have at least 1 IBD haplotypes from a segment of the chromosome which the length is *t* Morgan. This cumulative distribution function will be developed by the moment generating function.

Keywords: Full Sibs, the continuous time of Markov Chain, the moment generating function, and the cumulative distribution function.

### INTRODUCTION

5 The genes at a given locus of two related individuals are said to be identical by descent (IBD) if one is a 2 ysical copy of the other or both are physical copies of the same gene in a common ancestor, i. e. Both are copies that arose by DNA replication from the same ancestral sequence without any intervening mutation.

In 1983, Donnelly (Stefanov, 2002) calculates the probability that individuals in a given relationship share any part of the genome IBD by means of the continuous time of Markov Chain, whose states are the vertices of a hypercube. This research was considered by Stefanov (2000) as the first concerning IBD calculation, in the framework of the continuum model. Stefanov examine more deep than Donnelly and give to us a methodology to find of the cumu8 tive distribution function of the proportion of genome shared identical by descent on chromosome segments by two individuals 1 a grandparent-type relationship (2000) and a pair of full or half sibs (2002). Stefanov's methodology could be applied because we had data about distance between two locus for each chromosome. This data was produced by Human Genome Project (HGP). One of the HGP's goals was mapped and

sequenced human's genome. This mapping is very useful to know what kind of gene in human's chromosome and the expression from that gene. This paper adapted Markov Chain model from Stefanov (2002). Stefanov found the cumulative distribution function from the infimum of the length chromosome segment by the characteristic function, but in this paper we found cumulative distribution function by the moment generating function.

### FULL SIBS

cording to Stefanov (2002), for a pair of full sibs the relevant hypercube is four dimensional and the coordinates are either 0 or 1. 0 denote a grand paternal DNA and 1 denote a grand maternal DNA. The first two coordinates indicate the parental transmissions for sib one and the other two do the same for sib two. So, first and third coordinates denote DNA whiche were generated by their mother and second and fourth coordinates denote DNA was which were generated by their father. The DNA at a locus or a segment from a homologous chromosome is called haplotype. From that coding process we will get 16 states of the hypercube. The sixteen states of the hypercube can be divided into three groups of vertices indicating whether 0, 1, or 2 haplotypes are

 Ist International Conference on Actuarial Science and Statistics (ICASS 2014)

 AIP Conf. Proc.
 1692, 020007-1
 -020007-5; doi: 10.1063/1.4936435

 © 2015 AIP Publishing LLC 978-0-7354-1339-9/\$30.00

This article is copyrighted as indicated in the article. Reuse of AIP content 020007c1 to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 120.164.40.190 On: Thu. 24 Dec 2015 03:53:43

shared IBD at a locus. So, this case can be analyzed by a continuous time Markov Chain with three states and i, j = 0, 1, 2. The states are denoted by 0, 1, and 2 corresponding to the number of shared IBD haplotype. Cross over was occurred for this case can be looked at Figure 1.

Table 1. Relationship between vertices in hypercube with number of shared IBD haplotype							
vertice	number	vertices in	number of	vertices in	number of	vertices in	number
s in	of	hypercube	shared	hypercube	shared IBD	hypercube	of
hyperc	shared		IBD		haplotype		shared
ube	IBD		haplotype				IBD
	haplotyp						haplotyp
	e						e
11,1	2	01,11	1	10,01	0	01,00	1
1							
11,1	1	11,00	0	01,01	2	0 0, 1 0	1
0							
11,0	1	10,10	2	00,11	0	0 0, 0 1	1
1							
10,1	1	01,10	0	10,00	1	0 0, 0 0	2
1							



Figure 1.Sixteen possibility a copy gene at one lokus for a pair of full sibs.

In full sibs case, the continuous data on a chromosomal segment consists of the lengths of the consecutive pieces (subsegments) characterized by the number of haplotype shared IBD. Let t be the length (in Morgan) of a chromosome segment of interest. The sojourn time in state i (i = 0, 1, 2), within time interval of length t, is the length of genome whose each location has i haplotype shared IBD by two sibs on that segment. Such a genome will be called briefly a genome with i haplotype IBD. The sojourn time in state 2, within time interval of length t, is the length of genome whose each

location has 2 haplotype shared IBD by two sibs on that segment.

4 The number of entries to state i, within time interval of length t, is related to the number of distinct pieces (sub segment), of a chromosomal segment of length t, on each of which exactly i haplotype ar 4 hared IBD.

The number of entries from state  $\theta$  to state I, within time interval of length t, is related to number of sub segment which have at least 1 IBD haplotypes from a chromosomal segment which t4 length is t Morgan. The number of entries from state I to state 2, within time interval of length t, is related to number of sub

This article is copyrighted as indicated in the article. Reuse of AIP content 020007c2 to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 120.164.40.190 On: Thu, 24 Dec 2015 03:53:43

## segment which have 2 IBD haplotypes from a chromosomal segment which the length is t Morgan.

**6**  $N_{ij}(t)$  denotes the number of one step transition from state *i* to state *j*, *i*, *j* = 0, 1, 2, and A(t) denote the initial state.

- a.  $N_{I2}(t)$ , given the initial state is I, counts the number of sub segment which have 2 IBD haplotypes from a chromosomal segment which the length is t Morgan.
- b.  $N_{0l}(t)$ , given the initial state is 0, counts number of sub segment which have at least 1 IBD haplotypes from a chromosomal

segment which the length is t Morgan. For example 1 we were got  $N_{0l}(t) = 4$ .

Introduce the following stopping time, denoted  $\tau_{k, ij/l}$ , i. e.  $\tau_{k, ij/l} = inf \{t: N_{ij}(t) = k\}$ , k = 1, 2, ..., and  $\tau_{0, ij/l} = 0, i, j, l = 0, 1, 2$ , and l is initial state of the process.

 $H_2^{FS}(t)$  denotes the number of sub segment which have 2 IBD haplotypes from a chromosomal segment which the length is t Morgan and  $H_{1,2}^{FS}(t)$  denotes the number of sub segment which have at least 1 IBD haplotypes from a chromosomal segment which the length is t Morgan.

The relationship between  $\tau_{k,ij/l}$  and  $H_2^{FS}(t)$  and  $H_{l,2}^{FS}(t)$  as follow:

a.  $H_2^{FS}(t) > k \Leftrightarrow (\tau_{k+1, 12/0} \le t, A(0)=0)$  or  $(\tau_{k+1, 12/1} \le t, A(0)=1)$  or  $(\tau_{k, 12/0} \le t, A(0)=2)$ ; b.  $H_{1,2}^{FS}(t) > k \Leftrightarrow (\tau_{k+1, 01/0} \le t, A(0)=0)$  or  $(\tau_{k, 01/1} \le t, A(0)=1)$  or  $(\tau_{k, 01/2} \le t, A(0)=2)$ ; Consequence, cumulative distribution function of  $H_2^{FS}(t)$  and  $H_{1,2}^{FS}(t)$  related to cumulative distribution function of  $\tau_{k, ij/l}$  as follows:

$$P(H_{2}^{FS}(t) \le k) = I - \sum_{i=0}^{l} P(\tau_{k+1,12/i} \le t | A(0) = i) P(A(0) = i) - P(\tau_{k,12/i} \le t | A(0) = 2) P(A(0) = 2), k = 0, I, \dots \quad (1)$$

$$P(H_{I,2}^{FS}(t) \le k) = I - \sum_{i=1}^{2} P(\tau_{k,01/i} \le t | A(0) = i) P(A(0) = i) - P(\tau_{k+1,01/i} \le t | A(0) = 0) P(A(0) = 0), k = 0, I, \dots \quad (2)$$

$$Proposition 1:$$

So, if we want to find cumulative distribution function of  $H_2^{FS}(t)$  and  $H_{I_2}^{FS}(t)$  is needed conditional cumulative distribution function of  $\tau_{k,ij/I}$ . Proposition 1 and 2 give moment generating function for conditional cumulative distribution function of  $\tau_{k,ij/I}$ .

 $M_{\tau_{12/0}}(y) = M_{\tau_{12/2}}(y) = \frac{1}{\left\{2\left(1 - \frac{y}{4}\right)^2 - 1\right\}^k}, \quad k$ 

= 1, 2, .... Proof: (look Hongki, 2005).

#### **Proposition 2:**



Proof: (look Hongki, 2005).

So, if we want to find cumulative distribution function of  $H_2^{FS}(t)$  and  $H_{I_2}^{FS}(t)$ , we can be used inversion formula for Fourier Transformation. We can get the value of

 $F_{H_2^{FS}(t)}(k)$  dan  $F_{H_{1,2}^{FS}(t)}(k)$  for t = 0,5 M, t = 1,75 M, t = 1,80 M, t = 1,85 M, and t = 1,90 M with Maple programs.

This article is copyrighted as indicated in the article. Reuse of AIP conten 020007.3 to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 120.164.40.190 On: Thu, 24 Dec 2015 03:53:43

k	$F_{H_2^{FS}(0.5)}(k)$	$F_{H_2^{FS}(1.75)}(k)$	$F_{H_2^{FS}(1.80)}(k)$	$F_{H_2^{FS}(1.85)}(k)$	$F_{H_2^{FS}(1.90)}(k)$
0	0.4062679319	0.0937667072	0.0884317309	0.0834003023	0.0786551480
1	0.8588306426	0.3646658317	0.3501710083	0.3361450751	0.3225816053
2	0.9856720669	0.6842471426	0.6682925538	0.6523555050	0.6364619108
3	0.9992529136	0.8902490552	0.8806849178	0.8707699154	0.8605206715
4	0.9999769183	0.9727891255	0.9692222097	0.9653683926	0.9612223486
5	0.9999995336	0.9950446667	0.9941462791	0.9931314794	0.9919917599
6	0.9999999933	0.9993172633	0.9991553644	0.9989637240	0.9987385233
7	0.99999999999	0.9999268937	0.9999050922	0.9998780027	0.9998446382
8	1	0.9999937716	0.9999915021	0.9999885383	0.9999847074
9	1	0.9999995691	0.9999993815	0.9999991237	0.9999987736
10	1	0.9999999753	0.9999999628	0.9999999445	0.9999999185
11	1	0.9999999988	0.9999999981	0.9999999970	0.9999999954
12	1	1	0.99999999999	0.99999999999	0.9999999998
13	1	1	1	1	1

**Table 2.**  $F_{H_{2}^{ES}(t)}(k)$ , for t = 0,5 M, t = 1,75 M, t = 1,80 M, t = 1,85 M, and t = 1,90 M

**Table 3.**  $F_{H_{t,s}^{ES}(t)}(k)$ , for t = 0,5 M, t = 1,75 M, t = 1,80 M, t = 1,85 M, and t = 1,90 M

	,				
k	$F_{H_{1,2}^{FS}(0.5)}(k)$	$F_{H_{1,2}^{FS}(1.75)}(k)$	$F_{H_{1,2}^{FS}(1.80)}(k)$	$F_{_{H_{1,2}^{FS}(1.85)}}(k)$	$F_{H_{1,2}^{FS}(1.90)}(k)$
0	0.0338338209	0.0002279706	0.0001866465	0.0001528133	0.0001251129
1	0.744868221	0.1870774734	0.1764901686	0.1664949780	0.1570600704
2	0.9727930630	0.5422541900	0.5238518480	0.5057951721	0.4881031400
3	0.9985510707	0.8262400955	0.8127332598	0.7989158381	0.7848206818
4	0.9999547565	0.9542580148	0.9486365761	0.9426239930	0.9362206610
5	0.9999990802	0.9913202361	0.9898078433	0.9881127924	0.9862240363
6	0.9999999868	0.9987690969	0.9984847147	0.9981501667	0.9977594838
7	0.99999999999	0.9998654296	0.9998260140	0.9997772814	0.9997175629
8	1	0.9999937716	0.9999841704	0.9999787240	0.9999717133
9	1	0.9999995691	0.9999988338	0.9999983526	0.9999977016
10	1	0.9999999753	0.9999999292	0.9999998948	0.9999998456
11	1	0.9999999988	0.9999999964	0.9999999943	0.9999999914
12	1	1	0.99999999999	0.9999999998	0.99999999997
13	1	1	1	1	1

### CONCLUSION

1. The sixteen states of the hypercube can be divided into three groups of vertices indicating whether 0, 1, or 2 haplotypes are shared IBD at locus. So, we can be analyzed cross over what happen at a pair of full sibs with a continuous time Markov Chain with three states.

2. The relationship between cumulative distribution function of  $H_2^{FS}(t)$  and  $H_{I_z}^{FS}(t)$  with cumulative distribution function of  $\tau_{k,ij/l}$  as follows:

$$P\left(H_{2}^{FS}(t) \le k\right) = 1 - \sum_{i=0}^{1} P\left(\tau_{k+1,12/i} \le t \left| A(0) = i \right) P\left(A(0) = i \right) - P\left(\tau_{k,12/i} \le t \left| A(0) = 2 \right) P\left(A(0) = 2 \right), \ k = 0, \ 1, \dots P\left(H_{2}^{FS}(t) \le k\right) = 1 - \sum_{i=0}^{2} P\left(\tau_{k,01/i} \le t \left| A(0) = i \right) P\left(A(0) = i \right) - P\left(\tau_{k+1,01/i} \le t \left| A(0) = 0 \right) P\left(A(0) = 0 \right), \ k = 0, \ 1, \dots P\left(H_{2}^{FS}(t) \le k\right) = 1 - \sum_{i=0}^{2} P\left(\tau_{k,01/i} \le t \left| A(0) = i \right) P\left(A(0) = i \right) - P\left(\tau_{k+1,01/i} \le t \left| A(0) = 0 \right) P\left(A(0) = 0 \right), \ k = 0, \ 1, \dots P\left(H_{2}^{FS}(t) \le k\right) = 1 - \sum_{i=0}^{2} P\left(\tau_{k,01/i} \le t \left| A(0) = i \right) P\left(A(0) = i \right) - P\left(\tau_{k+1,01/i} \le t \left| A(0) = 0 \right) P\left(A(0) = 0 \right), \ k = 0, \ 1, \dots P\left(H_{2}^{FS}(t) \le t \right) = 1 - \sum_{i=0}^{2} P\left(\tau_{k,01/i} \le t \left| A(0) = i \right) P\left(A(0) = i \right) - P\left(\tau_{k+1,01/i} \le t \left| A(0) = 0 \right) P\left(A(0) = 0 \right) \right) P\left(A(0) = 0 \right) P\left(A(0)$$

This article is copyrighted as indicated in the article. Reuse of AIP content 020097:4to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 120.164.40.190 On: Thu, 24 Dec 2015 03:53:43

3. At a pair of full sibs, for  $k_{I} = k_{2}$ , but  $t_{I} < t_{2}$ , then  $F_{H_{2}^{FS}(t_{1})}(k_{1}) > F_{H_{2}^{FS}(t_{2})}(k_{2})$  and  $F_{H_{1,2}^{FS}(t_{1})}(k_{1}) > F_{H_{1,2}^{FS}(t_{2})}(k_{2})$ .

### REFERENCES

- Brown ,T. A. (1999). Genomes. New York: John Wiley & Sons.
- Donnelly, K. (1983). The Probability that Related Individuals Share Some Section of The Genome Identical by Descent. Theoretical Population Biology. 23: 34 – 64.
- Folland, G. B. (1992). Fourier Analysis and Its Applications. California: Brooks/Cole Publishing Company.
- 4. Hartwell, L. H. dkk. *Genetics: From Genes* to Genomes. New York: Mc Graw Hill.
- 5. Hongki, J. 2005. Distribusi Pewarisan Genom yang Identical by Descent pada Kasus Dua individu Bersaudara Kandung

(Full Sibs) dan Dua Individu Bersaudara Tiri (Half Sibs). Tesis S2.

- Loeve, M. (1963). *Probability Theory*. Toronto: D. Van Nostrand Company.
- Ross, S. M. (1996). Stochastic Processes. New York: John Wiley&Sons, Inc.
- 8. Strachan, T. (1992). *The Human Genome*.Manchester: Bios Scientific Publisher.
- Stefanov, V. T. (2000). Ditribution of Genome Shared Identical by Descent by Two Individuals in Grandparent Type Relationship. Genetics. 156: 1403 – 1410.
- Stefanov, V. T. (2002). Statistics on Continuous IBD Data: Exact Distribution Evaluation for a Pair of Full(Half) Sibs and a pair of a (Great-)grandchild with (Great-) grandparent. BMC Genetic 2002, 3: 7.
- 11. Suryo. (2003). Genetika Manusia. Yogyakarta: Gadjah Mada University Press.
- 12. Walpole, R. E, Myers, R. H. (1995). *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan*. Bandung: ITB.
- 13. Yatim, W. (2003). Genetika. Bandung:Penerbit Tarsito.

This article is copyrighted as indicated in the article. Reuse of AIP conten 020007.5 to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 120.164.40.190 On: Thu, 24 Dec 2015 03:53:43

# The distribution of genome shared identical by descent for a pair of full sibs by means of the continuous time Markov chain

ORIGIN	IALITY REPORT				
SIMIL	<b>%</b> ARITY INDEX	<b>5%</b> INTERNET SOURCES	5% PUBLICATIONS	<b>0%</b> STUDENT PAP	PERS
PRIMA	RY SOURCES				
1	Valeri T of gene segmer generat 2016 Publication	. Stefanov. "Dist tic material from it surviving to th ion", Journal of <i>J</i>	ribution of the n a chromosor e following Applied Proba	e amount nal bility,	1 %
2	WWW.CO	ursehero.com			1%
3	doaj.or	ce			1%
4	"CS2-PX Publication	S-21_HR.pdf", A	ctEd		1%
5	WWW.pi	cb.ac.cn			1%
6	<b>oldwww</b> Internet Sour	<b>/.ippt.pan.pl</b>			1%
7	Henry Ju "Analys	unus Wattimane is of the distribu	la, G. Haumah tion types of c	iu. Jepth	1%

and magnitude of tectonic earthquake 2018 in

## Lombok Island based on clusterization results using K-Means method", AIP Publishing, 2021

Publication

8	www.cambridge.org	<b>1</b> %
9	dokumen.pub Internet Source	<1 %
10	export.arxiv.org	<1 %
11	<b>nozdr.ru</b> Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		