Quantitative Methoden in der historisch-vergleichenden Linguistik

Michael Cysouw

cysouw@lmu.de
web.me.com/cysouw

Data

What kind of data is used for the language comparison ?

- Traditional trinity of language description
  - Grammar
  - Texts
  - Dictionary
- Generalized dictionary (“constructicon”)
  - Language particular constructions
  - Combination of phonemic structure and meaning/context

Language Comparison

What is the reason for a particular similarity between two languages?

- Historical reasons
  - Shared descent
  - Contact
- A-historical reasons
  - Inherent (“universal”) preference
  - Chance

Quantitative Approach

Why should linguists use quantitative methods ?

- Upscaling
  - dealing with more data
- Methodological innovation
  - Using ‘cumbersome’ methods
- Greater precision
  - Beyond discrete ‘yes/no’ results

History of Quantification

- Early quantitative historical analysis
  - Kroeber, & Chrétien (1937, 1939), Chrétien (1943), Kroeber (1960)
- Distribution of reflexes of reconstructions
  - Ross (1950), Davies & Ross (1975)
  - Holm (2000, 2003, 2007)

History of Quantification

- Swadesh-list approach
  - Swadesh (1952), Lees (1953), Merwe (1966)
  - Chretien (1962)
- (silent) further development
  - Dyen et al. (1967), Dyen (1992)
  - Sankoff (1970, 1972)
  - Dobson et al. (1972), Dobson (1978)
  - Black (1976)
  - Embleton (1986, 1991)

History of Quantification

- Chance of cognation
  - Justeson & Stephens (1979)
  - Ringe (1992)
- Russian developments
  - Dolgoposky (1986 [1964])
  - Yakhontov, Starostin

History of Quantification

- Dialectometry
  - Séguy (1973)
  - Goebl (1984, 2006), Embleton (1985)
  - Nerbonne et al (1999), Heeringa (2004), Heeringa et al. (2005, 2006)
- String Alignment
  - Covington (1996, 1998, 2004)
  - Kondrak (2002, 2003), Kondrak & Sherif (2006), Wesley & Kondrak (2005), Bergsma & Kondrak (2007a, 2007b)
  - Rødseth & Sellars (2006)

History of Quantification

- "New synthesis" with bioinformatics
  - Renfrew *et al.* (2000)
  - Warnow *et al.* (1996), Warnow (1997), Ringe *et al.* (1998, 2002), Nakhleh *et al.* (2005a, 2005b)
  - Pagel (2000), Gray & Jordan (2000), Gray & Atkinson (2003), Atkinson & Gray (2005), Atkinson *et al.* (2005, 2008), Pagel *et al.* (2007)
  - Lohr (1999), Haggerty (2000a, 2000b), McMahon & McMahon (2003, 2005), McMahon et al. (2005)
  - Rexova (2002, 2006)
  - Kessler (2001, 2005, 2006, 2007)
  - Brown *et al.* (2008), Holman *et al.* (2008)

Comparison Biology-Linguistics

- Platnick & Cameron (1977)
- Atkinson & Gray (2005)
- Holm (2007)

“go” periphrasis in Tsezic

IDS Main Page
Simple Browsing
Advanced Browsing
Download Data
Technical Background

The Intercontinental Dictionary Series

Founding Editor:
Mary Ritchie Key (University of California, Irvine)
General Editor:
Bernard Comrie (Max Planck Institute for Evolutionary Anthropology, Leipzig)

Purpose: The purpose of the IDS is to establish a database where lexical material across the continents is organized in such a way that comparisons can be made. Historical studies, comparative, and theoretical linguistic research can be based on this documentation. This is a long term cooperative project that will go on for the next generation or so and will involve linguists all over the world. It is aimed towards international understanding and cooperation. This is a pioneering effort that will have global impact. The purpose also contributes to preserving information on the little-known and "non-prestigious" languages of the world, many of which are becoming extinct.

Rationale: Information on languages of the world is scattered over all the continents and islands and published in dozens of languages and scripts. There is need of a database where one can find comparable material to formulate hypotheses and test and validate those theories. For example, theories on intercontinental connections have been proposed on the basis of the distribution of 'sweet potato' and yet there is no single source, where words with this meaning can be found in many languages. Good quantitative and statistical studies are almost impossible to do now in non-Western languages. The IDS will provide a quantitative base for a scientific approach to language analysis and comparisons. The IDS will provide the research tools necessary for expanding studies such as phonological theory, word formation, language change, lexical distribution, symbolism and onomatopoeia, classification, and other ideas that have to do with history of people and migrations. The IDS will serve not only as a synonym dictionary but as an index to meaning and to cultures of various people around the earth.

Plan of Series: The IDS series may appear as 1) a volume with 25 or more languages recorded; 2) a fascicle with 5 to 10 languages recorded; 3) in single WordLists, which are archived until enough are gathered to make up a fascicle or volume. A list of fascicles and volumes in progress is available from the general editor.

Procedure: The IDS is developed in cooperation and complementation with other research projects. Throughout the world there are linguistic activities from establishing of databases in universities and think-tanks to publishing grammar series and literacy materials, to individual projects such as the Tibetan dictionary project. Many projects seek to make linguistic data accessible in a format that will allow generalizations to be made. The computer now gives us the potential for tying together linguistic databases. The IDS editors will continue to monitor linguistic activity around the world, both for choosing the languages for forthcoming compilations and for collaboration with other research teams.

http://lingweb.eva.mpg.de/ids/

IDS number	English	Hinukh
2.330	marry	xoddoho yiƛ’a
10.470	go	iƛ’a
10.471	climb	šid iƛ’a
10.472	go up	toššid iƛ’a
10.473	go down	ƛ’iyo iƛ’a
10.520	follow	hezzo iƛ’a
10.530	pursue	xalbun hezzo iƛ’a
10.650	drive	mašinaƛ’o iƛ’a
10.660	ride	guluƛ’o iƛ’a
18.210	speak	aƛ’i iƛ’a
20.430	attack	kʷerer biƛ’a
20.460	surrender	kʷar biƛ’a

	bezhta	hinukh	hunzib	khvarshi inxokvari	khvarshi khvarshi	tsez
burn	0	0	0	0	0	0
marry	1	1	0	0	0	1
yawn	0	0	0	0	1	0
shiver	0	0	0	0	1	0
move	1	0	1	0	0	0
fall	0	1	0	0	0	0
swim	0	0	0	0	0	0
creep	0	0	0	0	0	1
crouch	0	0	0	0	0	0
walk	1	0	1	0	0	0
go	1	1	1	1	1	1
go up	1	1	1	1	0	1
climb	0	1	1	0	0	0
go down	1	1	1	0	0	1
come	0	0	0	0	0	0
come back	0	0	0	0	0	0
go away	0	1	1	1	1	1
follow	1	1	0	0	0	1
pursue	0	1	0	0	0	0
arrive	0	0	0	1	0	0
drive	1	1	1	1	1	1
ride	1	1	1	1	1	1
lose	0	1	0	0	0	0
age	0	0	0	0	1	0
speak	0	1	0	0	0	0
attack	0	1	0	0	0	0
retreat	0	0	0	0	0	0
surrender	0	1	1	1	0	1
hunt	1	0	1	0	0	0
miss	1	0	0	0	0	0

hinukh

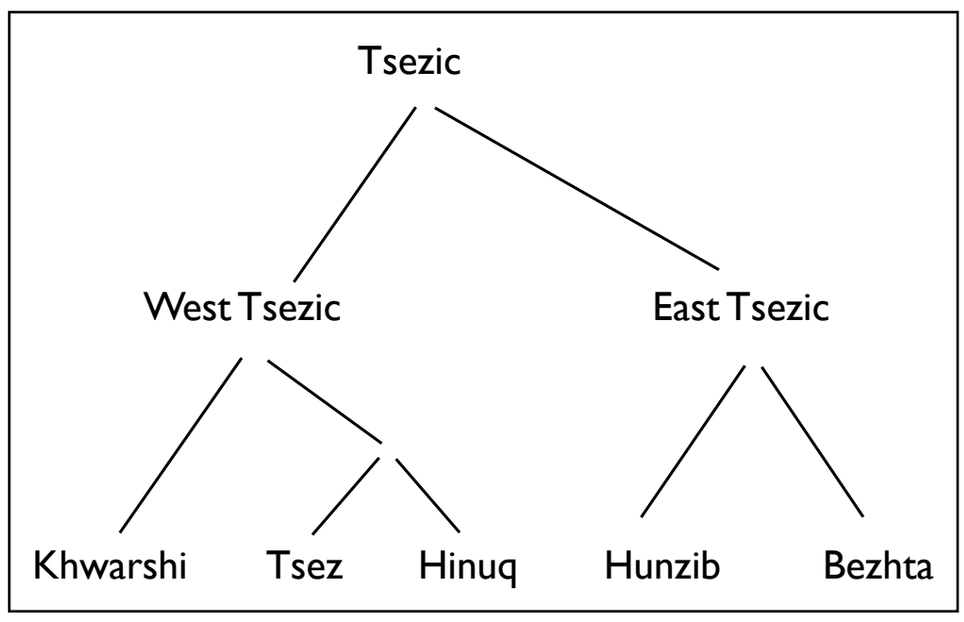
bezhta

hunzib

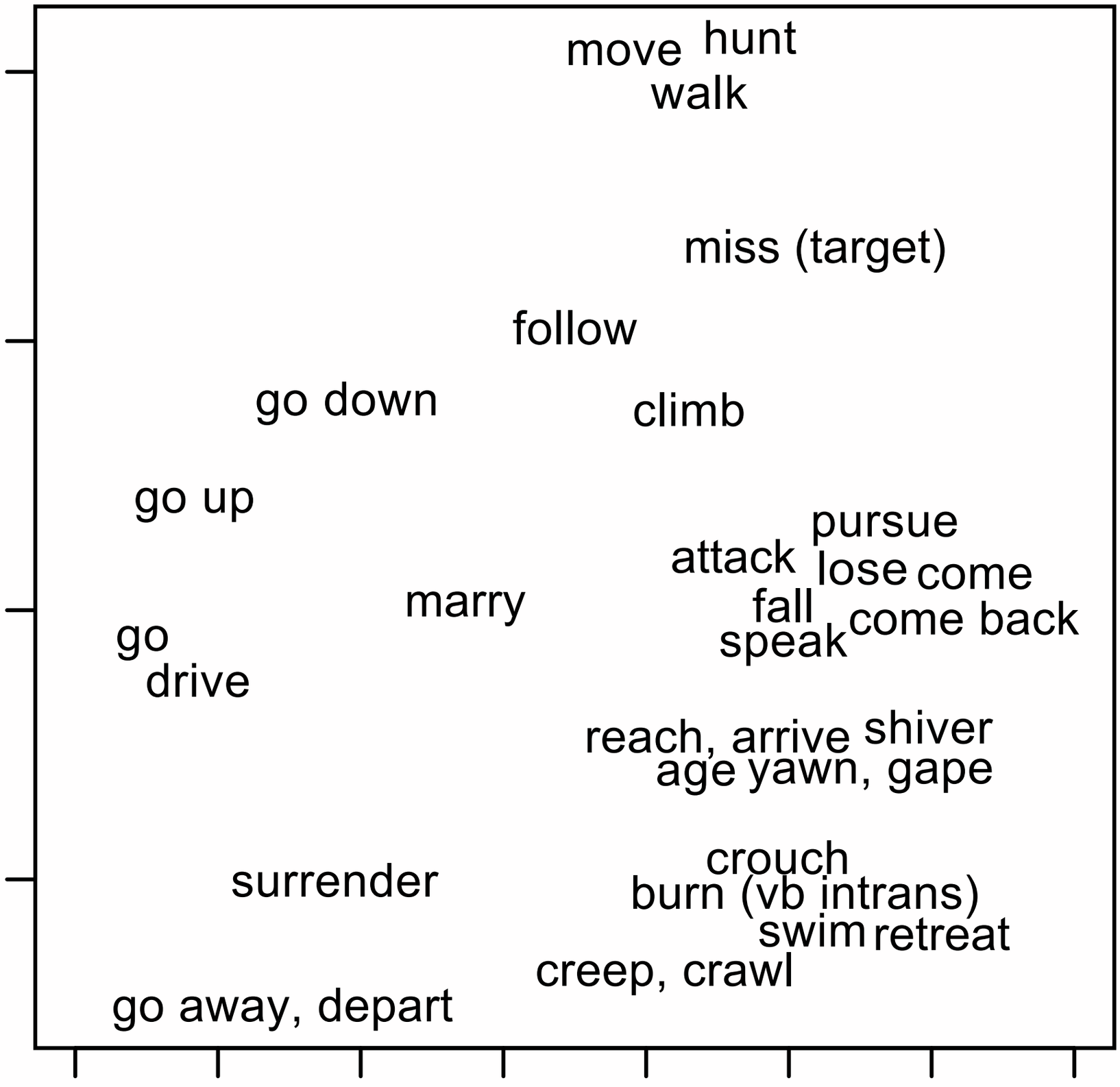
tsez

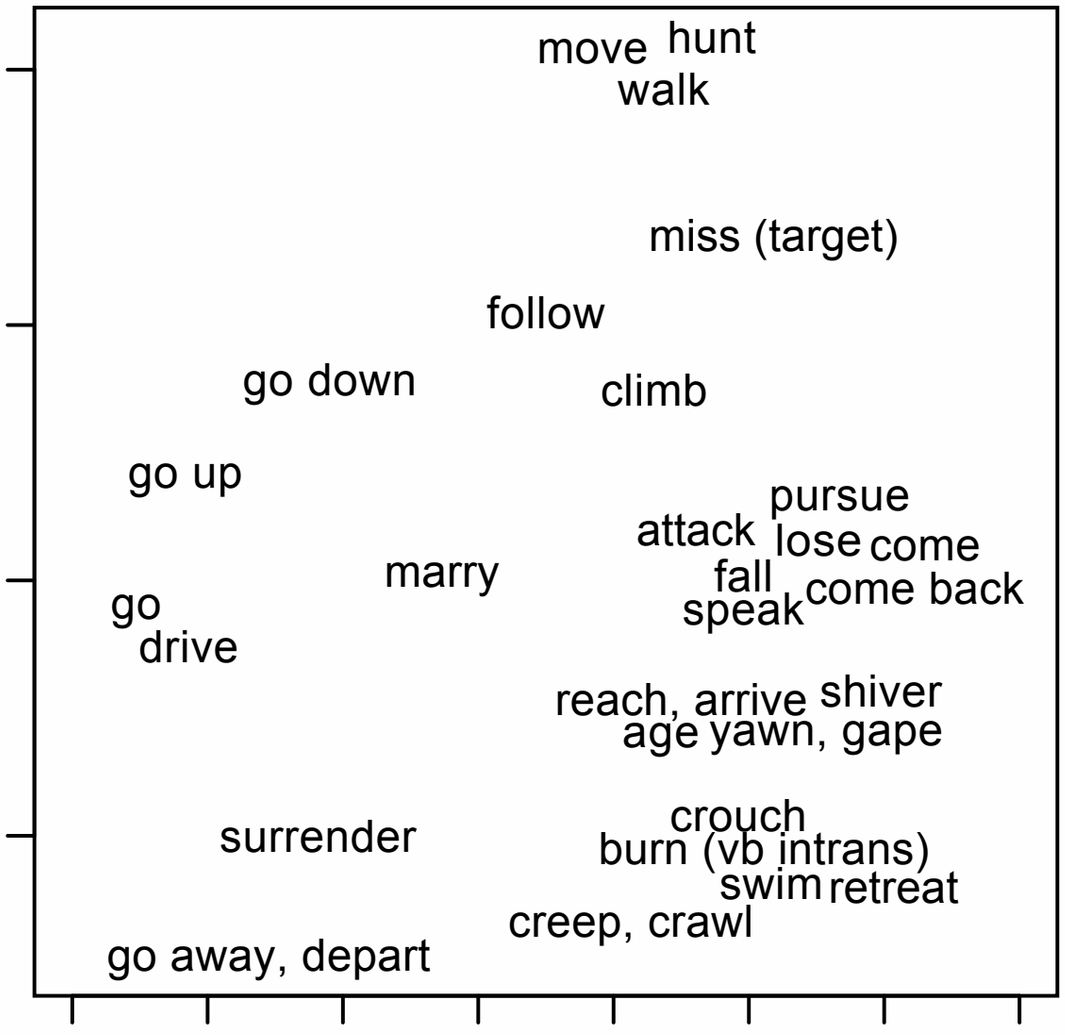
khvarshi inxokvari

khvarshi khvarshi

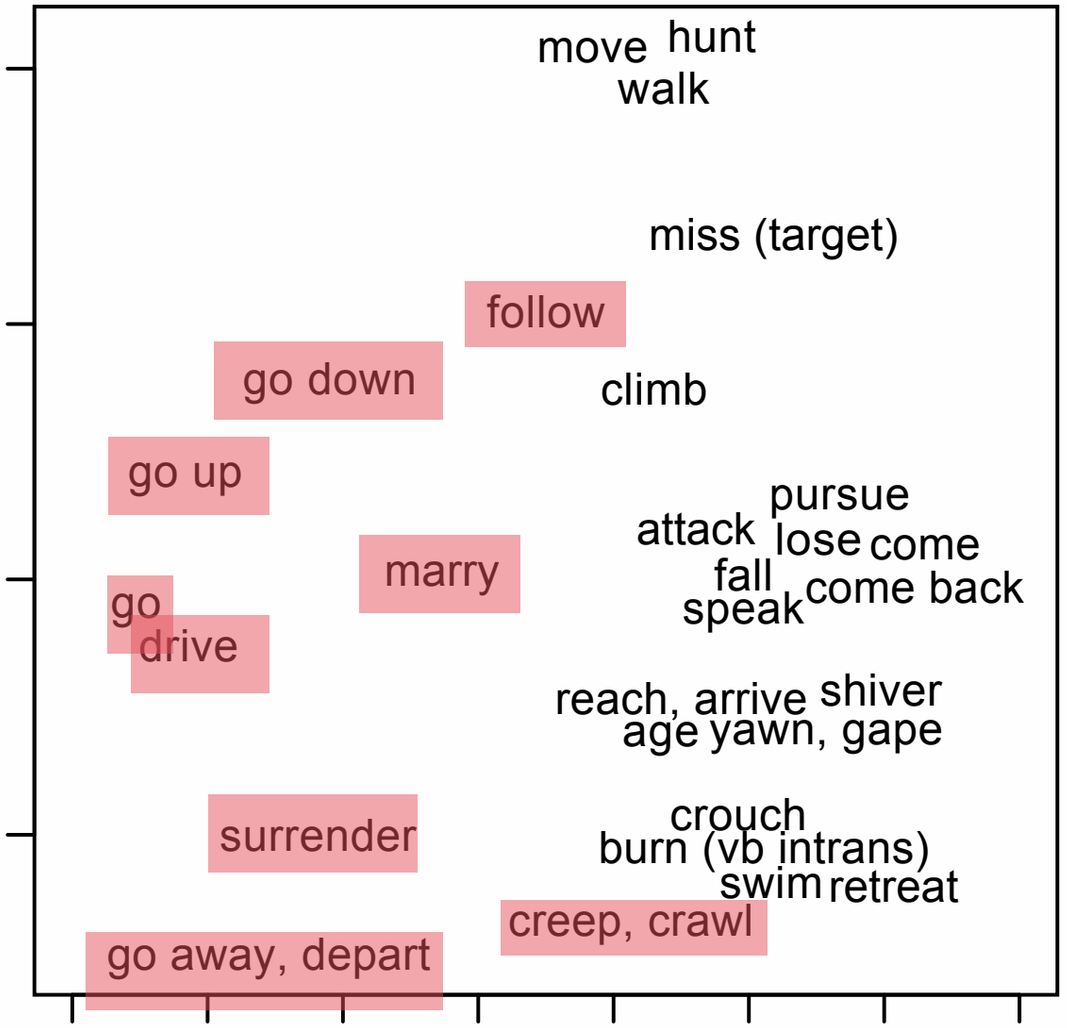


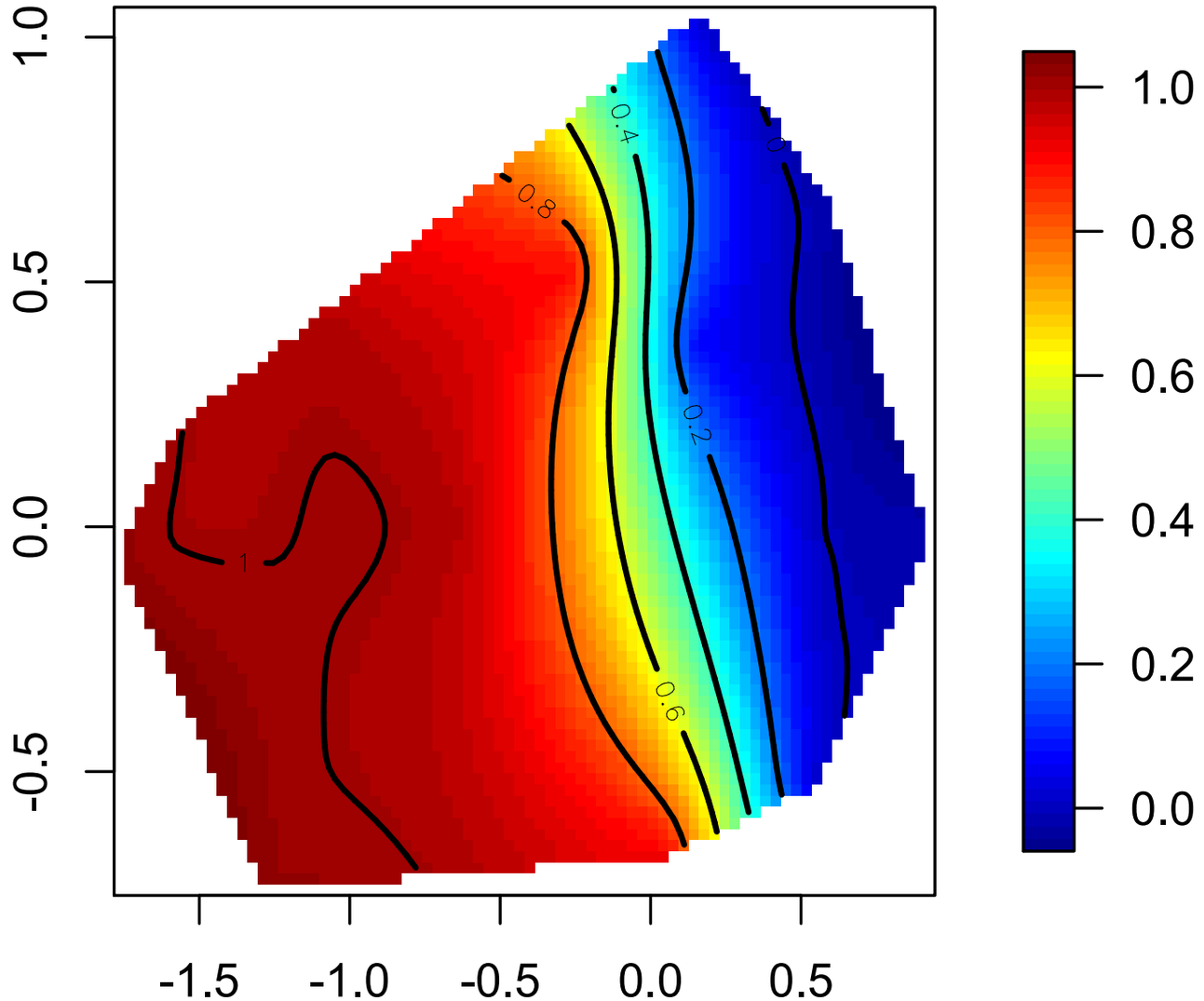
	bezhta	hinukh	hunzib	khvarshi inxokvari	khvarshi khvarshi	tsez
burn	0	0	0	0	0	0
marry	1	1	0	0	0	1
yawn	0	0	0	0	1	0
shiver	0	0	0	0	1	0
move	1	0	1	0	0	0
fall	0	1	0	0	0	0
swim	0	0	0	0	0	0
creep	0	0	0	0	0	1
crouch	0	0	0	0	0	0
walk	1	0	1	0	0	0
go	1	1	1	1	1	1
go up	**1**	**1**	**1**	**1**	**0**	**1**
climb	0	1	1	0	0	0
go down	**1**	**1**	**1**	**0**	**0**	**1**
come	0	0	0	0	0	0
come back	0	0	0	0	0	0
go away	0	1	1	1	1	1
follow	1	1	0	0	0	1
pursue	0	1	0	0	0	0
arrive	0	0	0	1	0	0
drive	1	1	1	1	1	1
ride	1	1	1	1	1	1
lose	0	1	0	0	0	0
age	0	0	0	0	1	0
speak	0	1	0	0	0	0
attack	0	1	0	0	0	0
retreat	0	0	0	0	0	0
surrender	0	1	1	1	0	1
hunt	1	0	1	0	0	0
miss	1	0	0	0	0	0


move
hunt
walk
miss (target)
follow
go down
climb
go up
pursue
attack
lose
come
marry
fall
come back
go
speak
drive
reach, arrive
shiver
age
yawn, gape
crouch
surrender
burn (vb intrans)
swim
retreat
creep, crawl
go away, depart

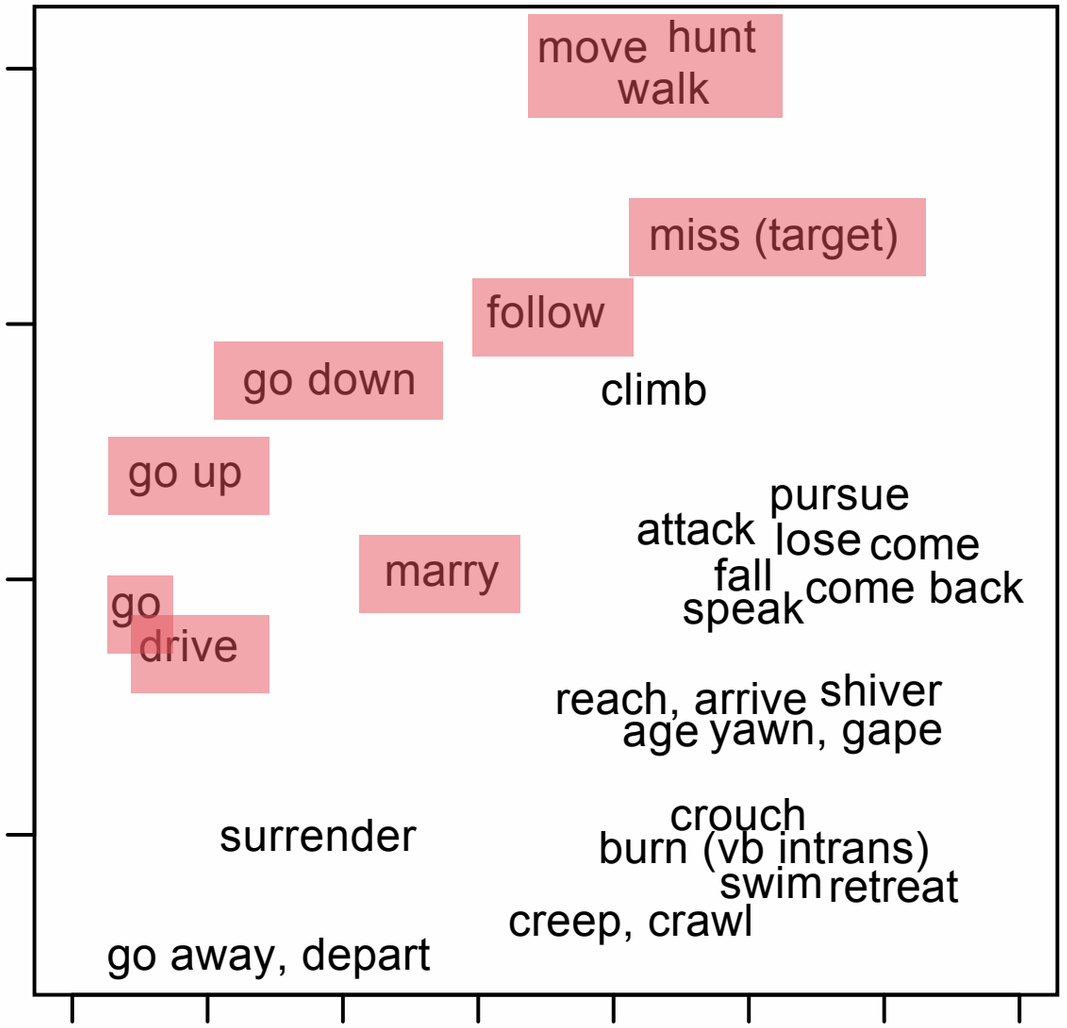

move
hunt
walk
miss (target)
follow
go down
climb
go up
pursue
attack
lose
come
marry
fall
come back
speak
go
drive
reach, arrive
shiver
age
yawn, gape
crouch
surrender
burn (vb intrans)
swim
retreat
creep, crawl
go away, depart

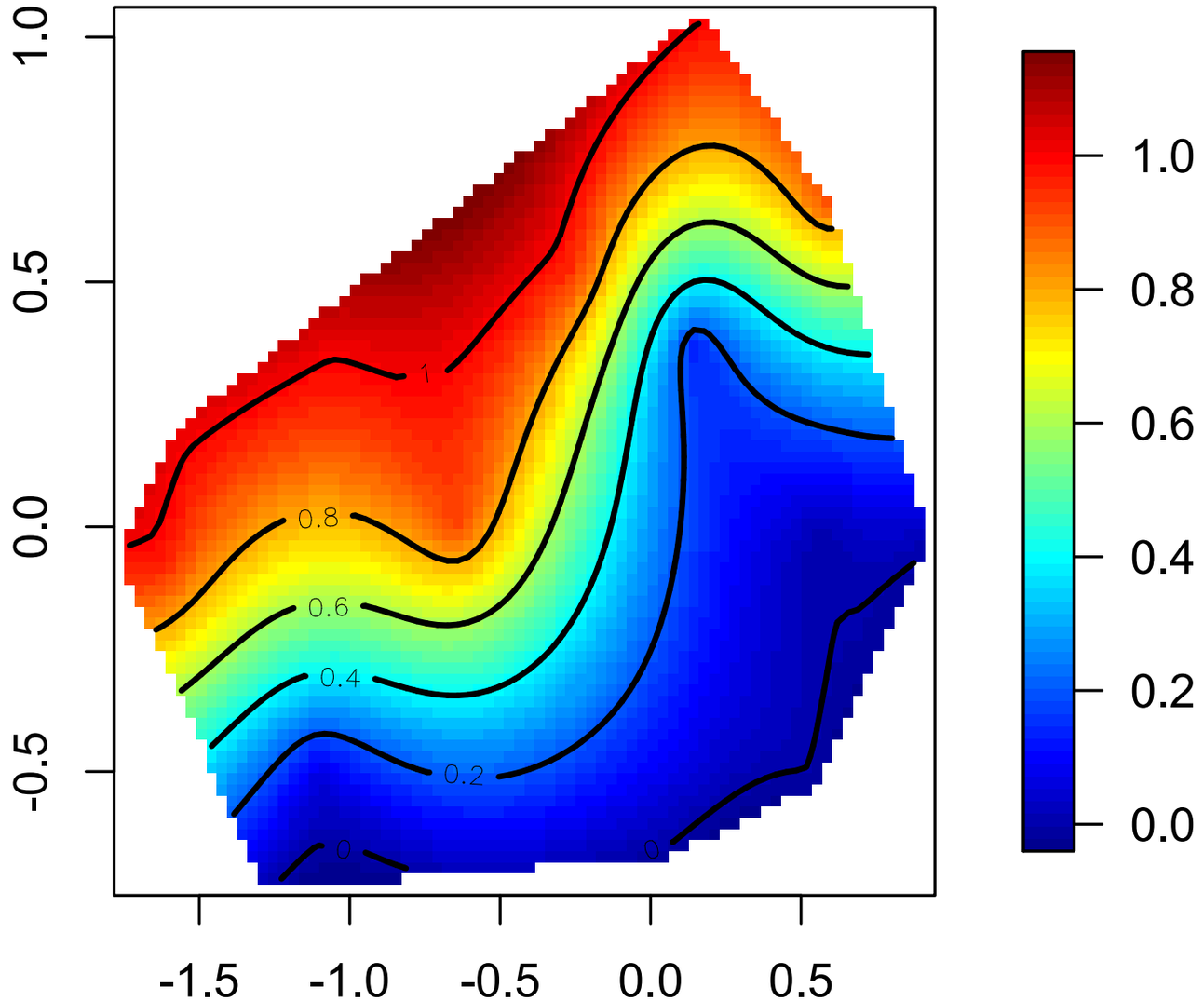
Tsez

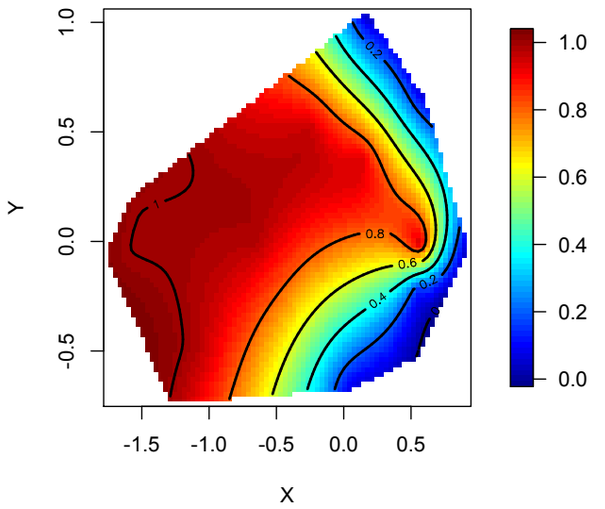

move
hunt
walk
miss (target)
follow
go down
climb
go up
pursue
attack
lose
come
marry
fall
come back
speak
go
drive
reach, arrive
shiver
age
yawn, gape
surrender
crouch
burn (vb intrans)
swim
retreat
creep, crawl
go away, depart

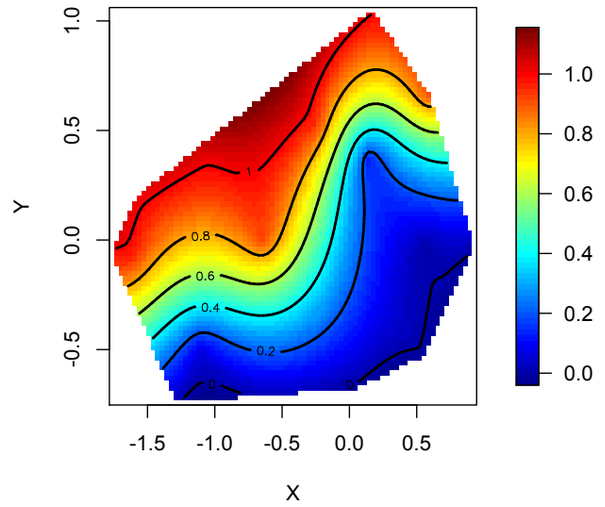

1.0
0.5
0.0
-0.5
Y
-1.5
-1.0
-0.5
0.0
0.5
0.4
0.8
0.2
0.6
1
1.0
0.8
0.6
0.4
0.2
0.0

Bezhta


move
hunt
walk
miss (target)
follow
go down
climb
go up
pursue
attack
lose
come
marry
fall
come back
speak
go
drive
reach, arrive
shiver
age
yawn, gape
crouch
surrender
burn (vb intrans)
swim
retreat
creep, crawl
go away, depart

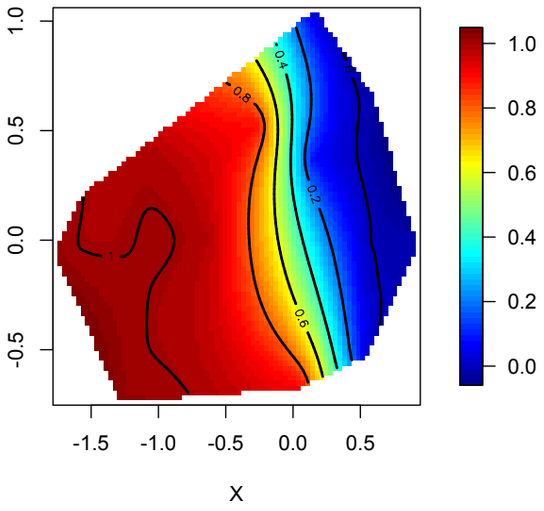

1.0
0.5
0.0
-0.5
Y
1
0.8
0.6
0.4
0.2
0
-1.5
-1.0
-0.5
0.0
0.5
1.0
0.8
0.6
0.4
0.2
0.0



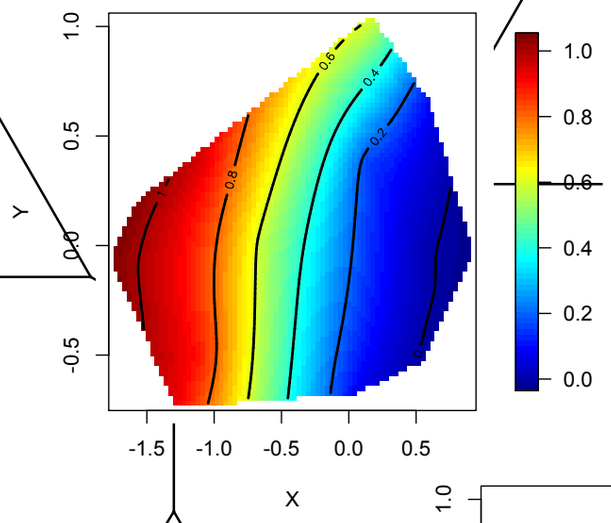


hinukh

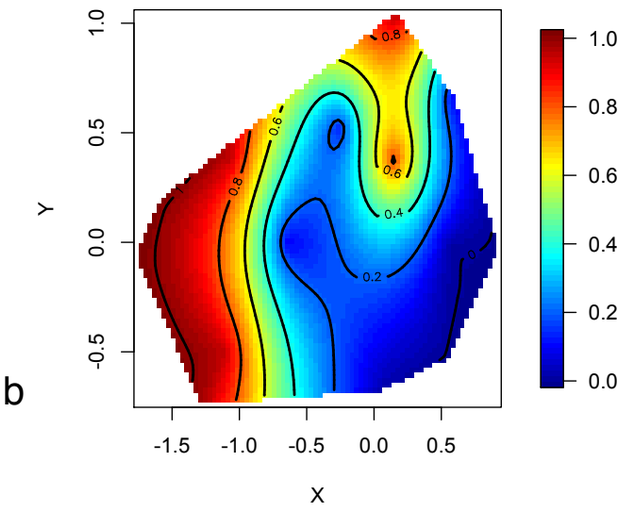
bezhta



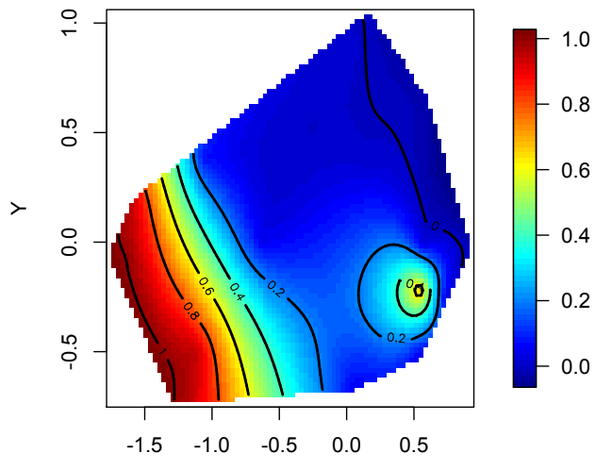
tsez



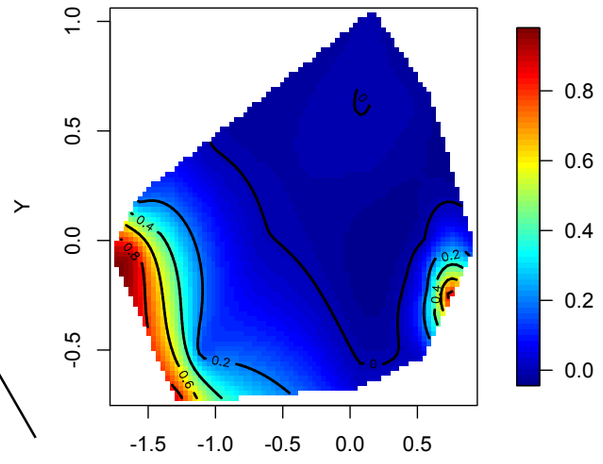
hunzib

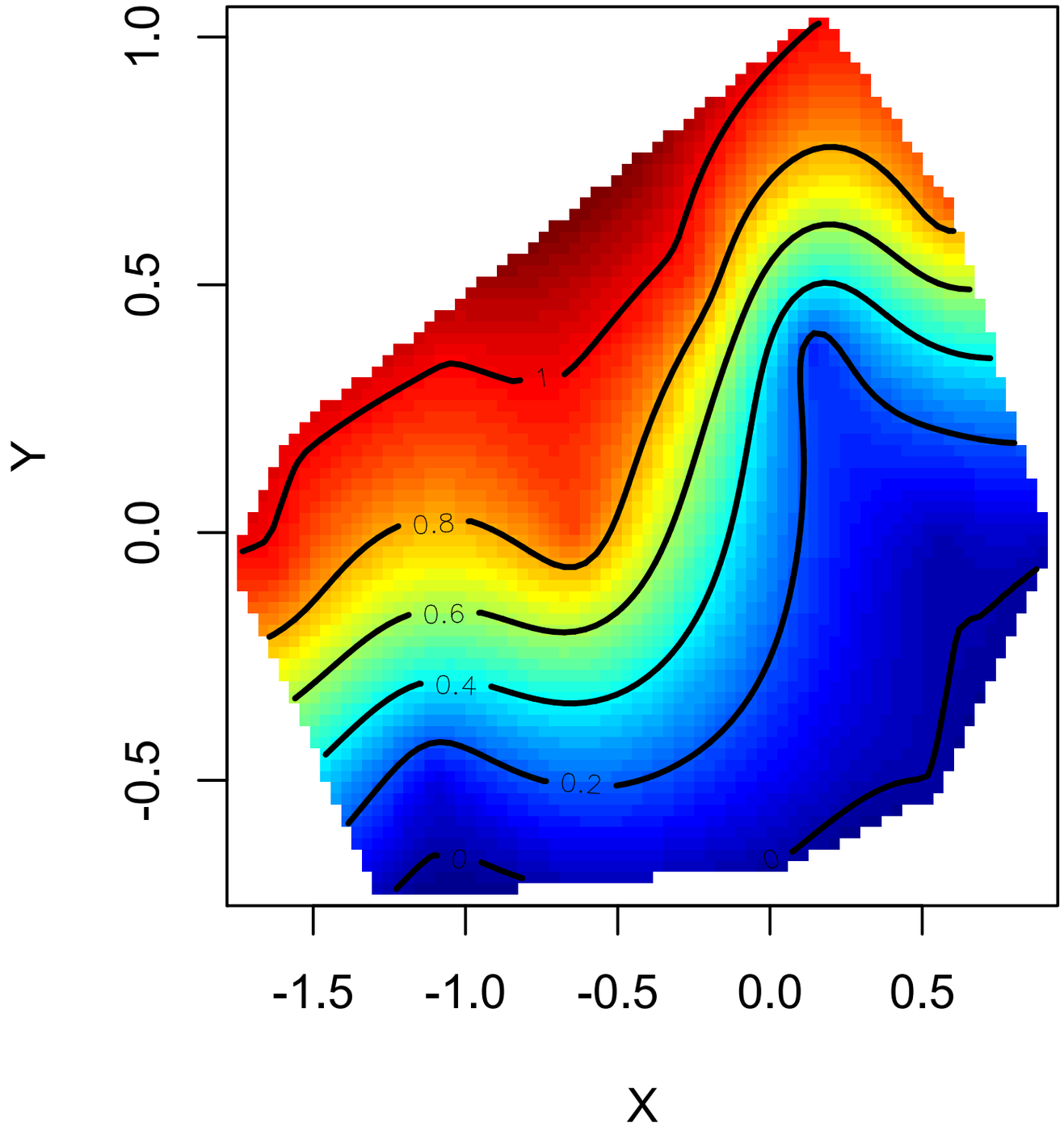


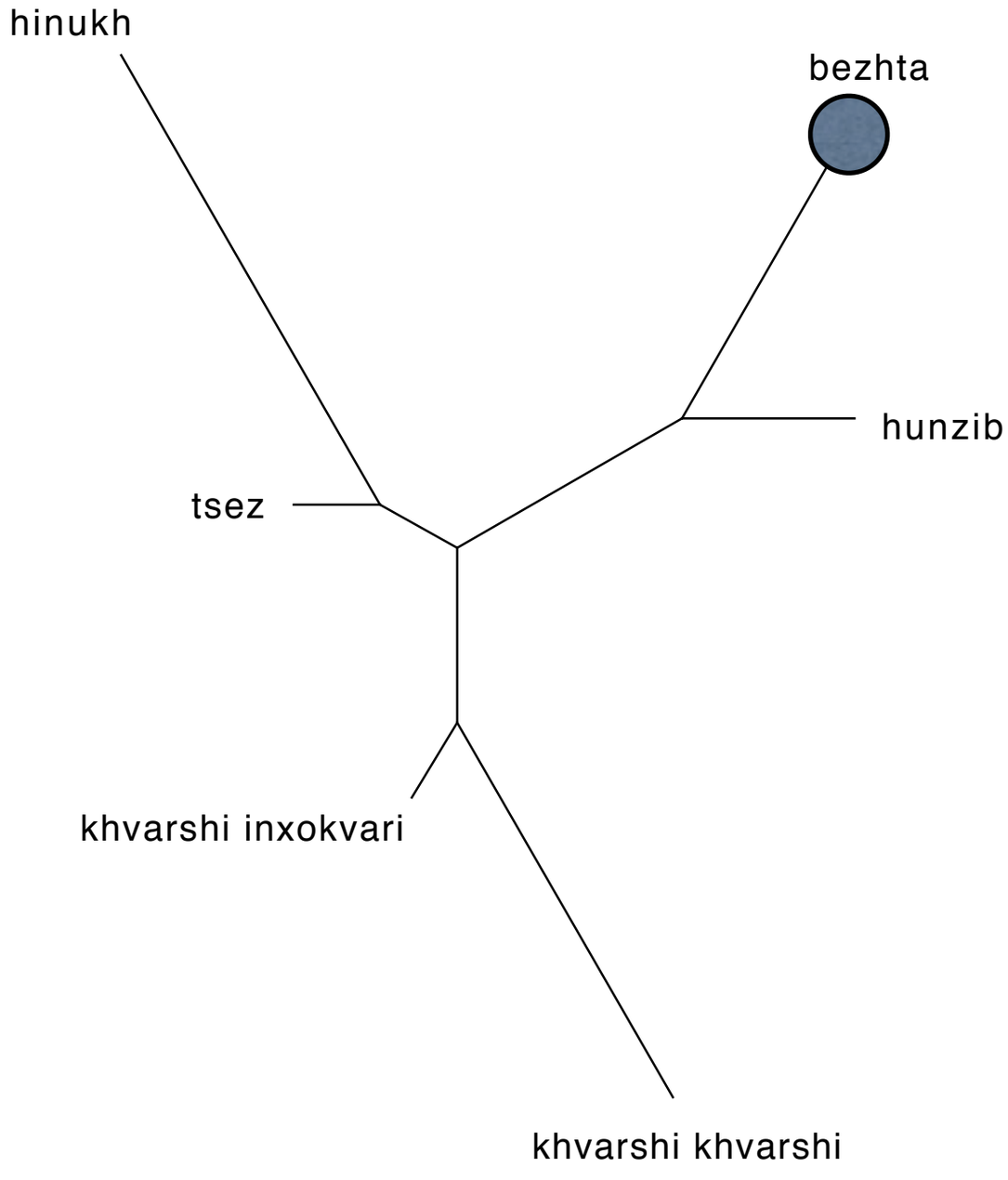
khvarshi inxokvari

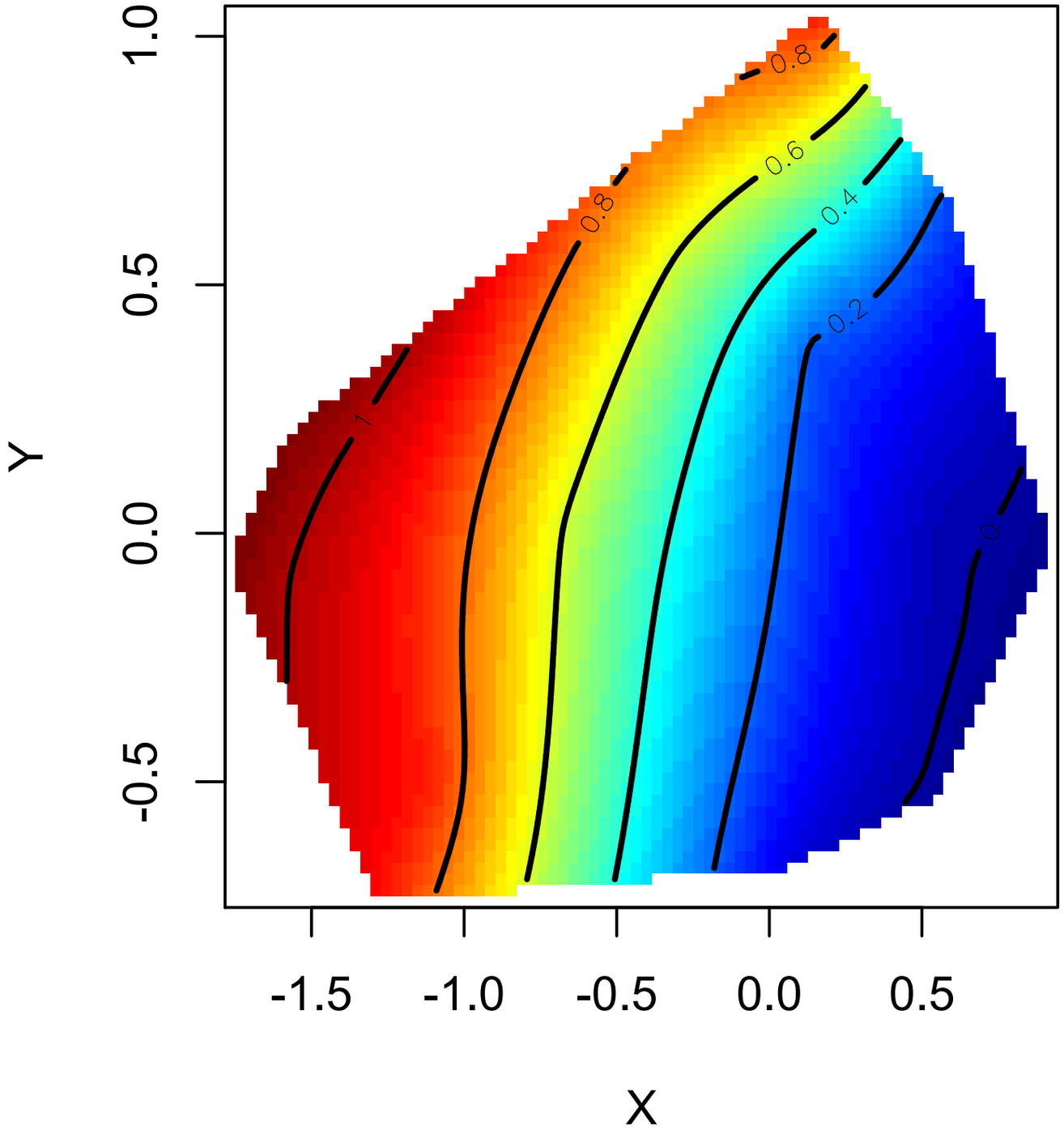


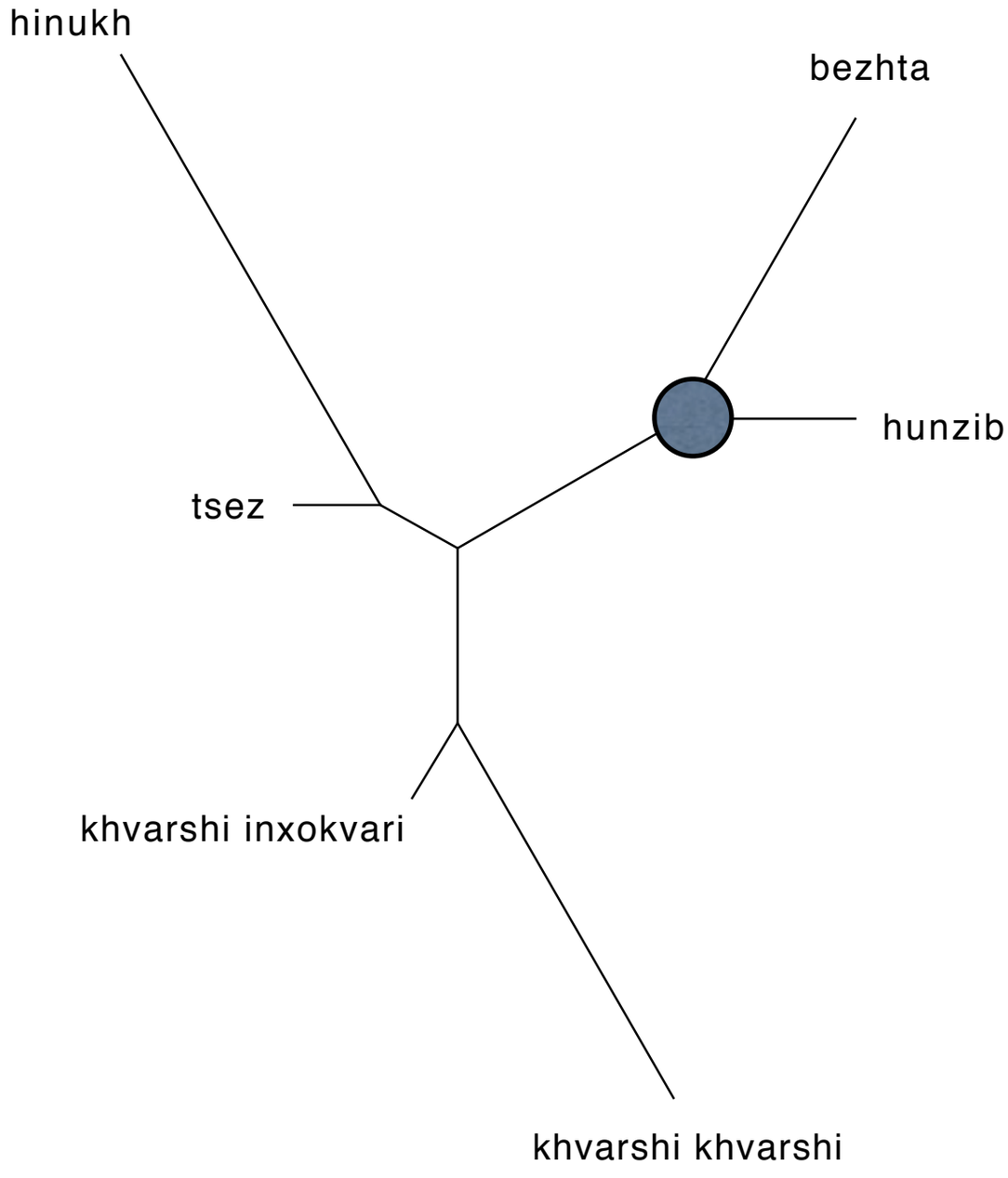
khvarshi khvarshi

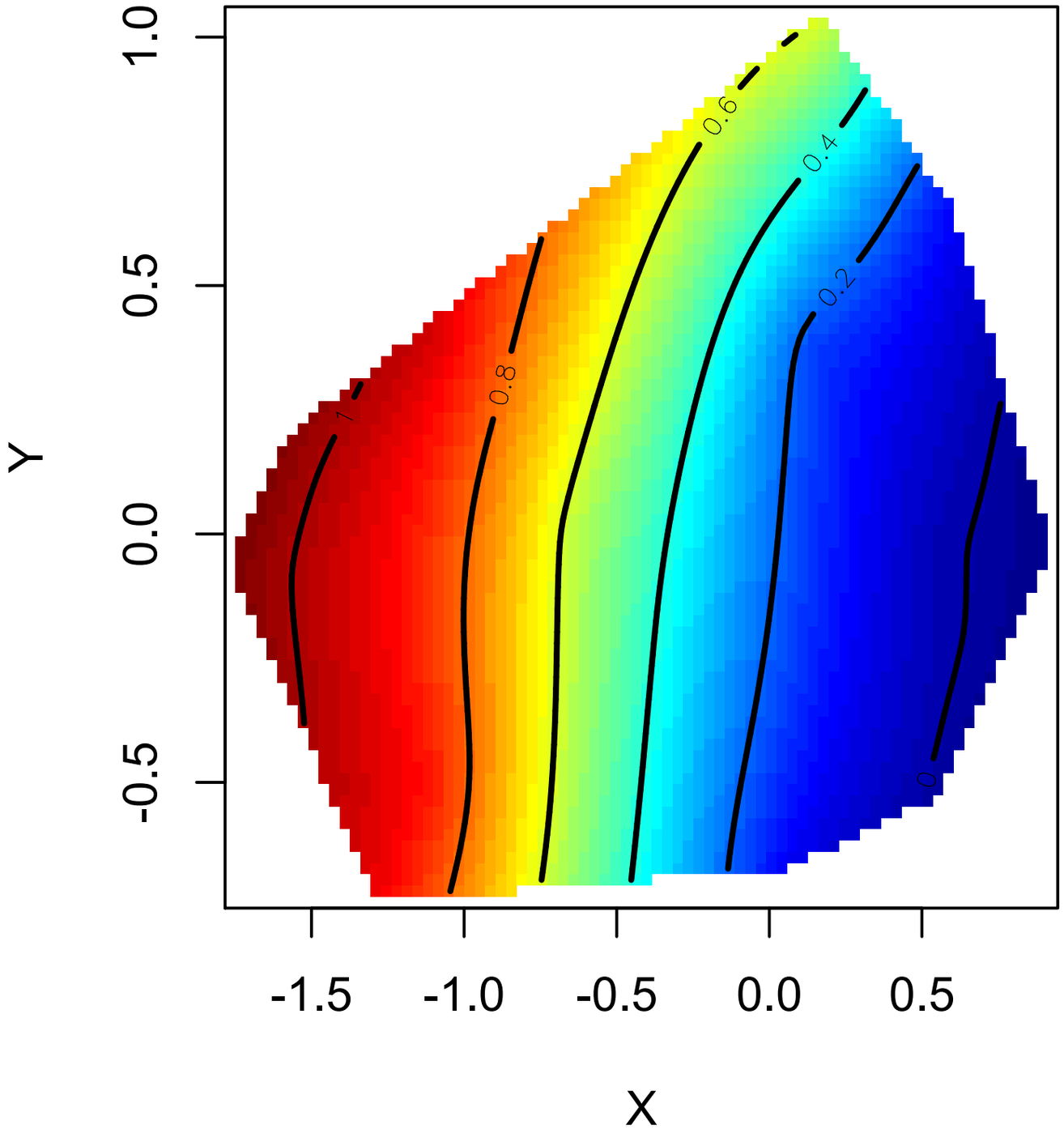


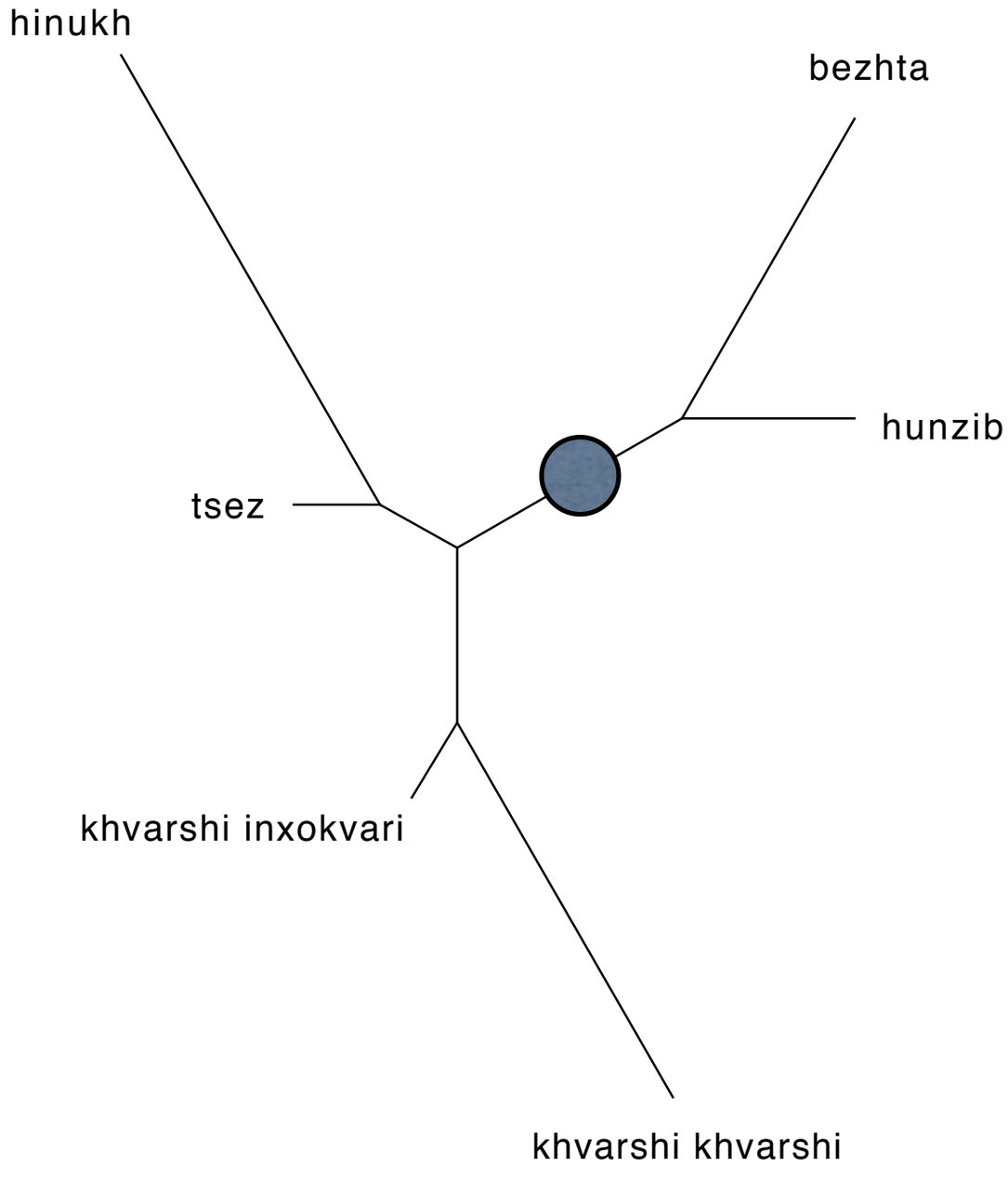

1.0
0.5
0.0
-0.5
Y
1
0.8
0.6
0.4
0.2
0
0
-1.5
-1.0
-0.5
0.0
0.5
X

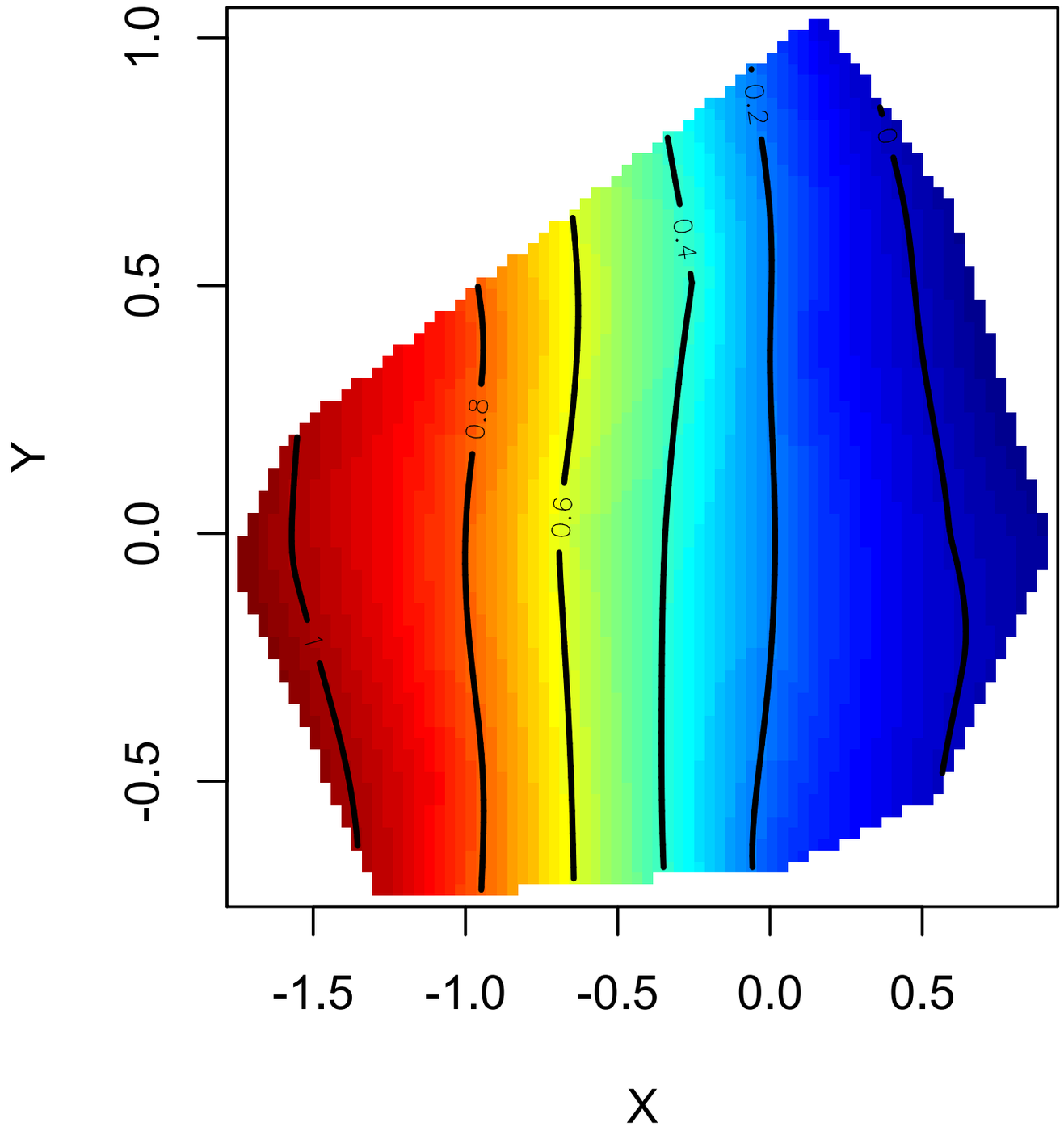

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

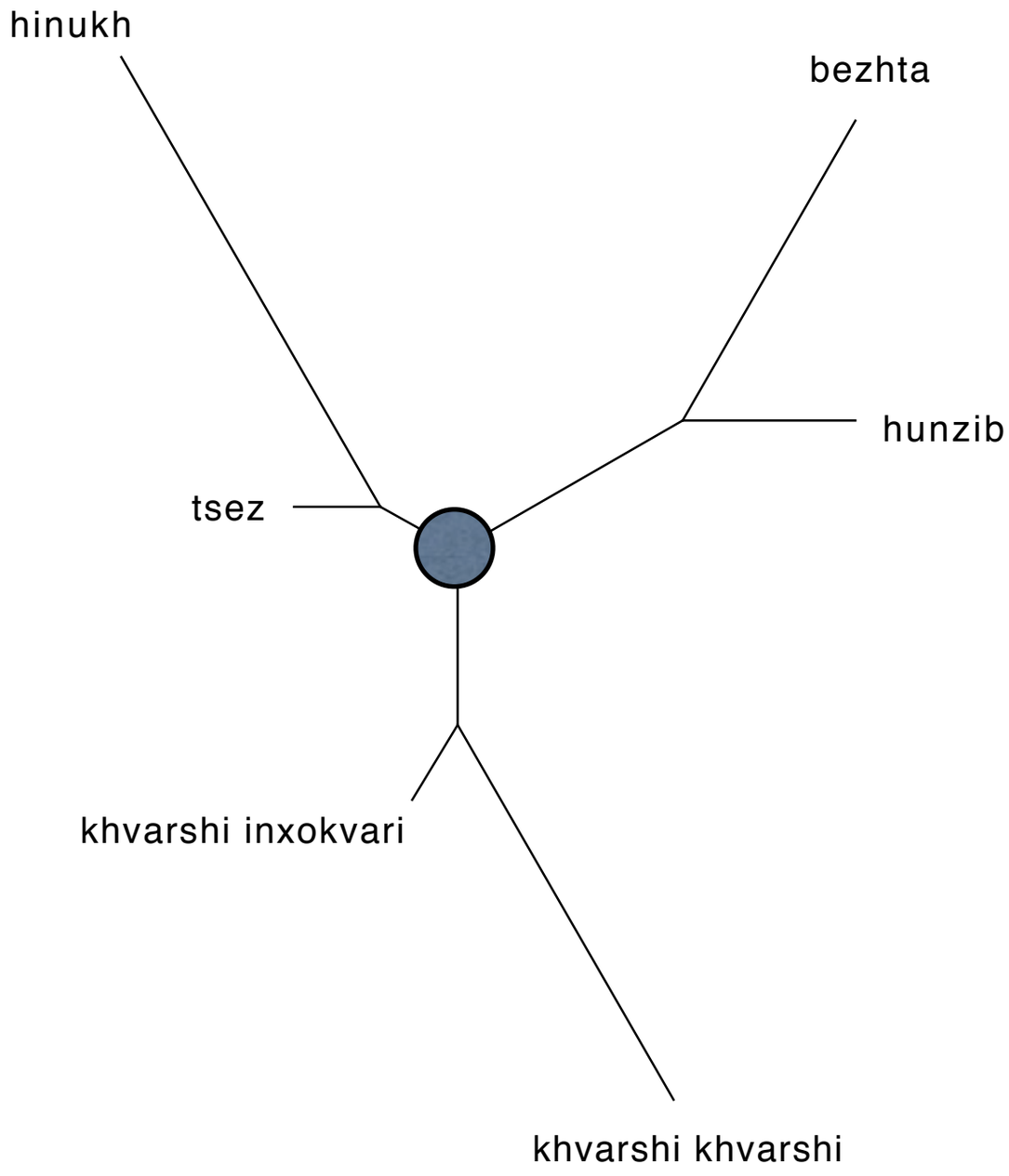

1.0
0.5
0.0
-0.5
Y
0.8
0.6
0.4
0.2
1
0
-1.5
-1.0
-0.5
0.0
0.5
X

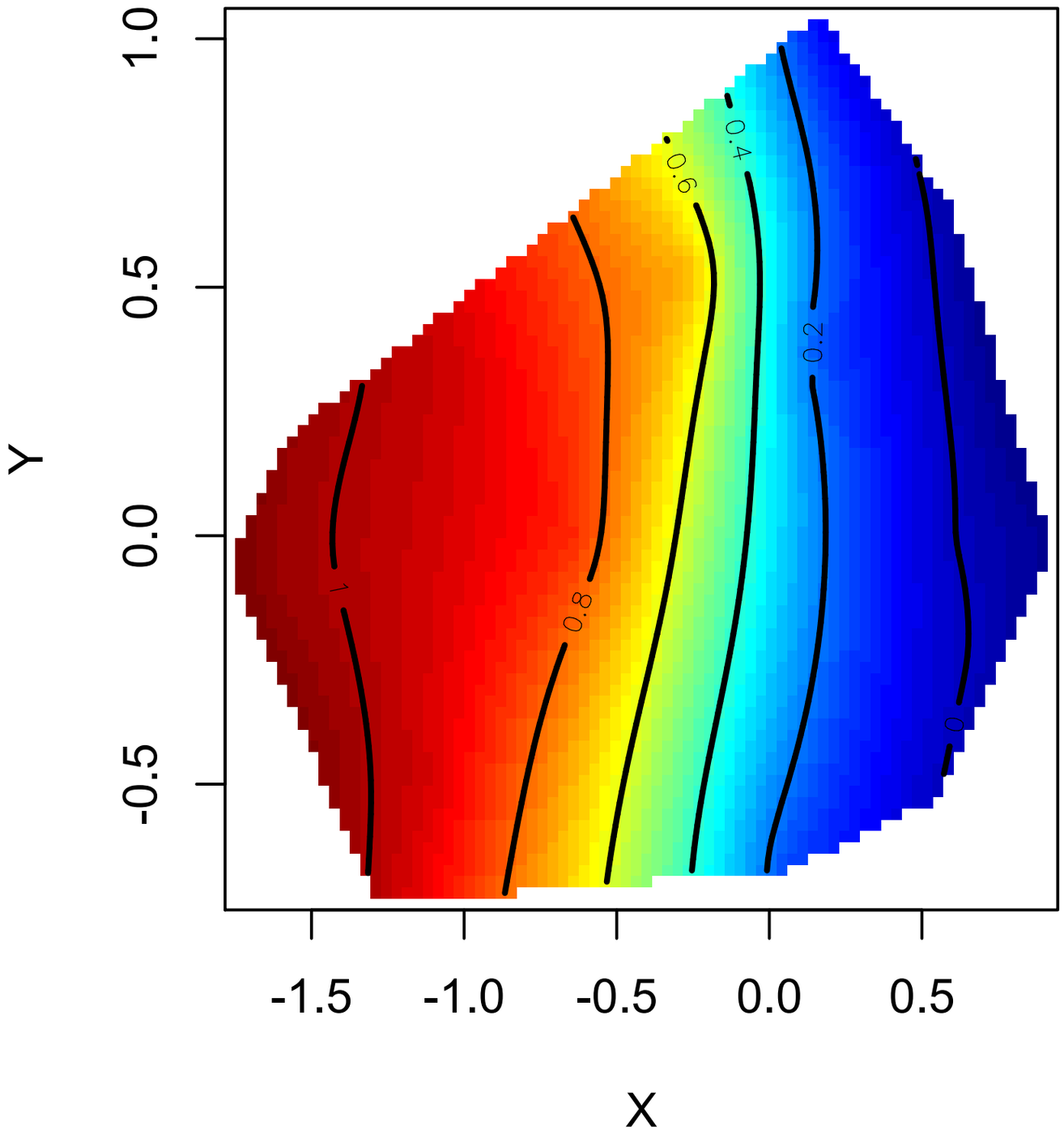

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

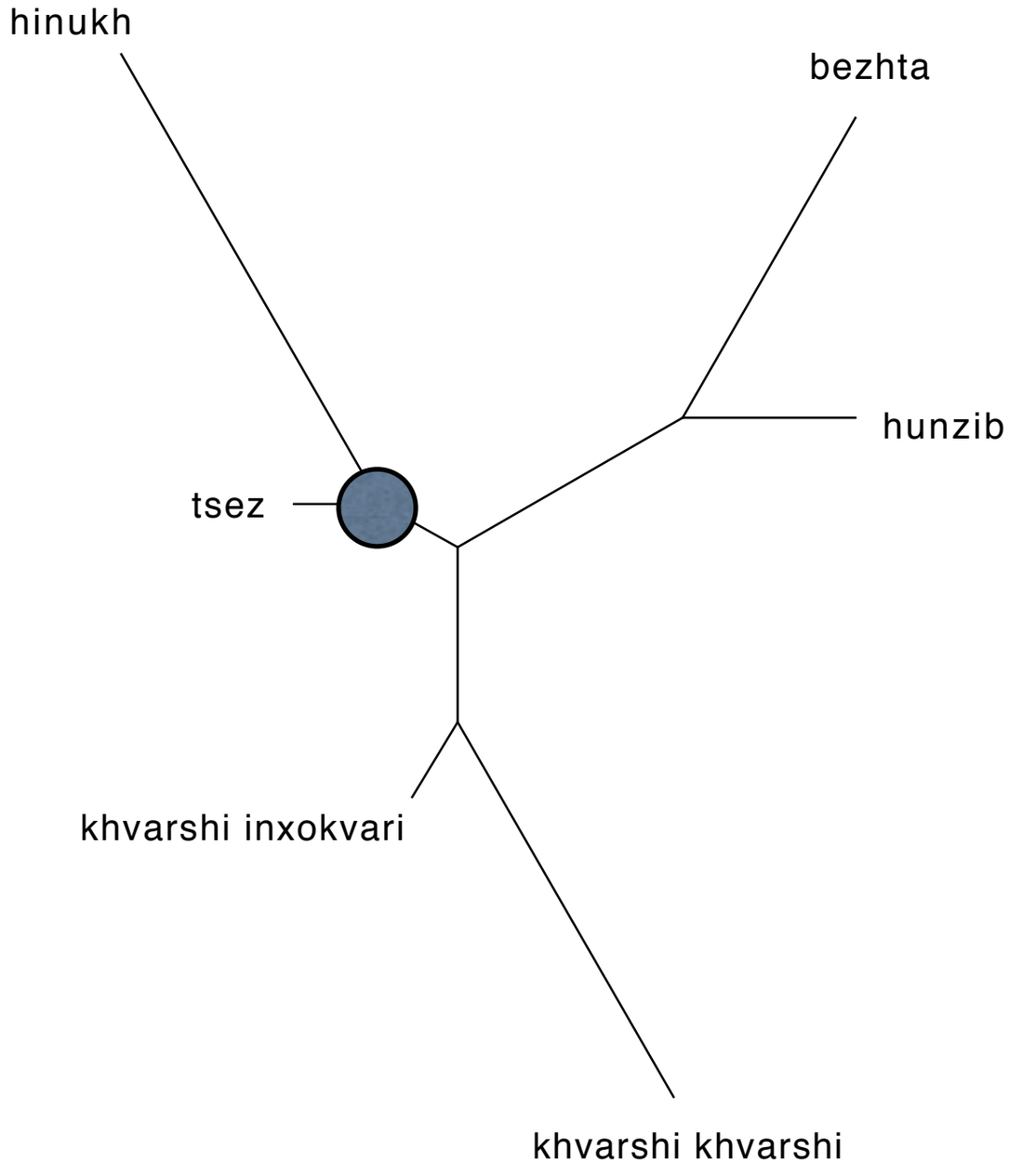

1.0
0.5
0.0
-0.5
Y
-1.5
-1.0
-0.5
0.0
0.5
X
1
0.8
0.6
0.4
0.2
0

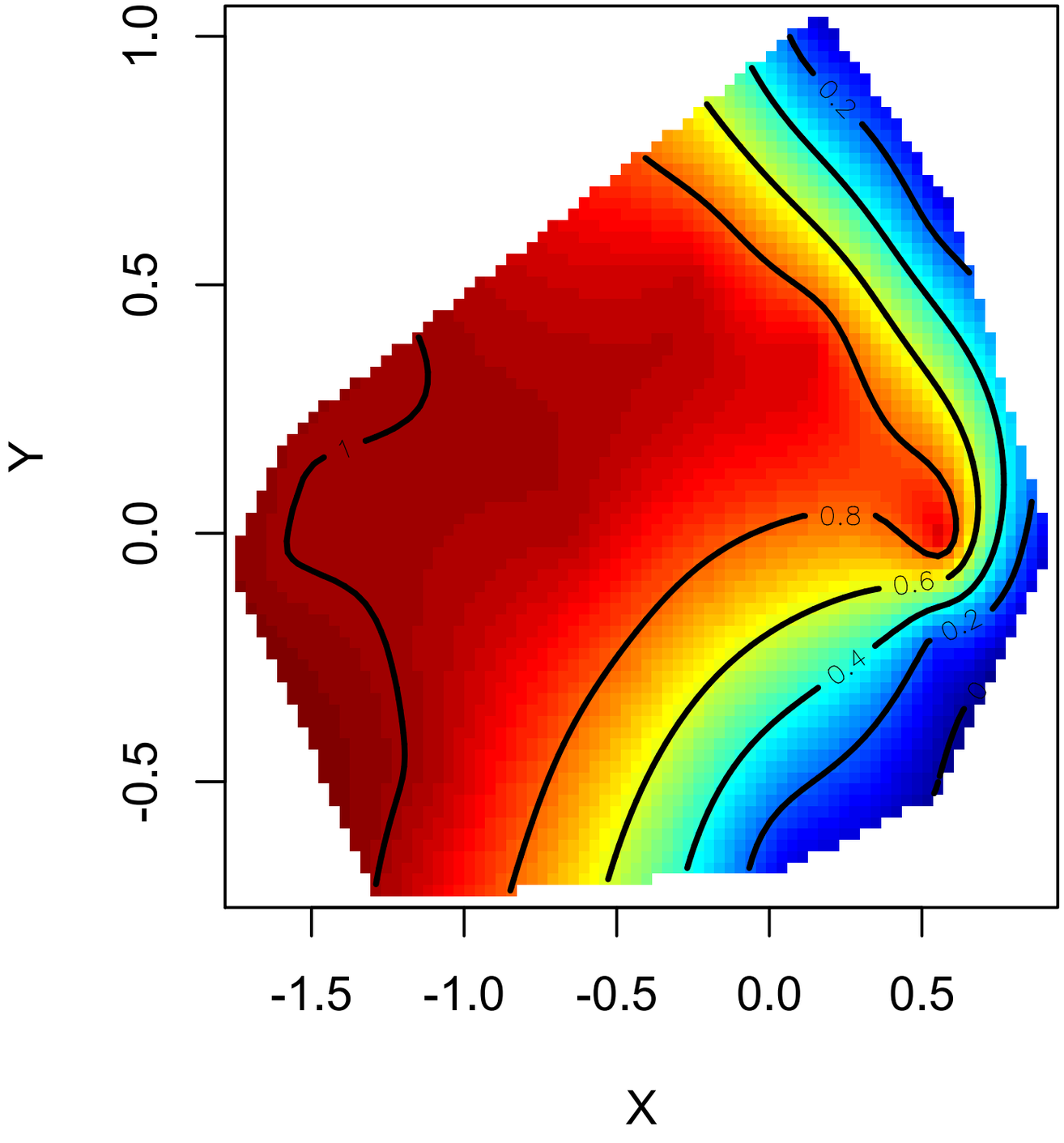

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

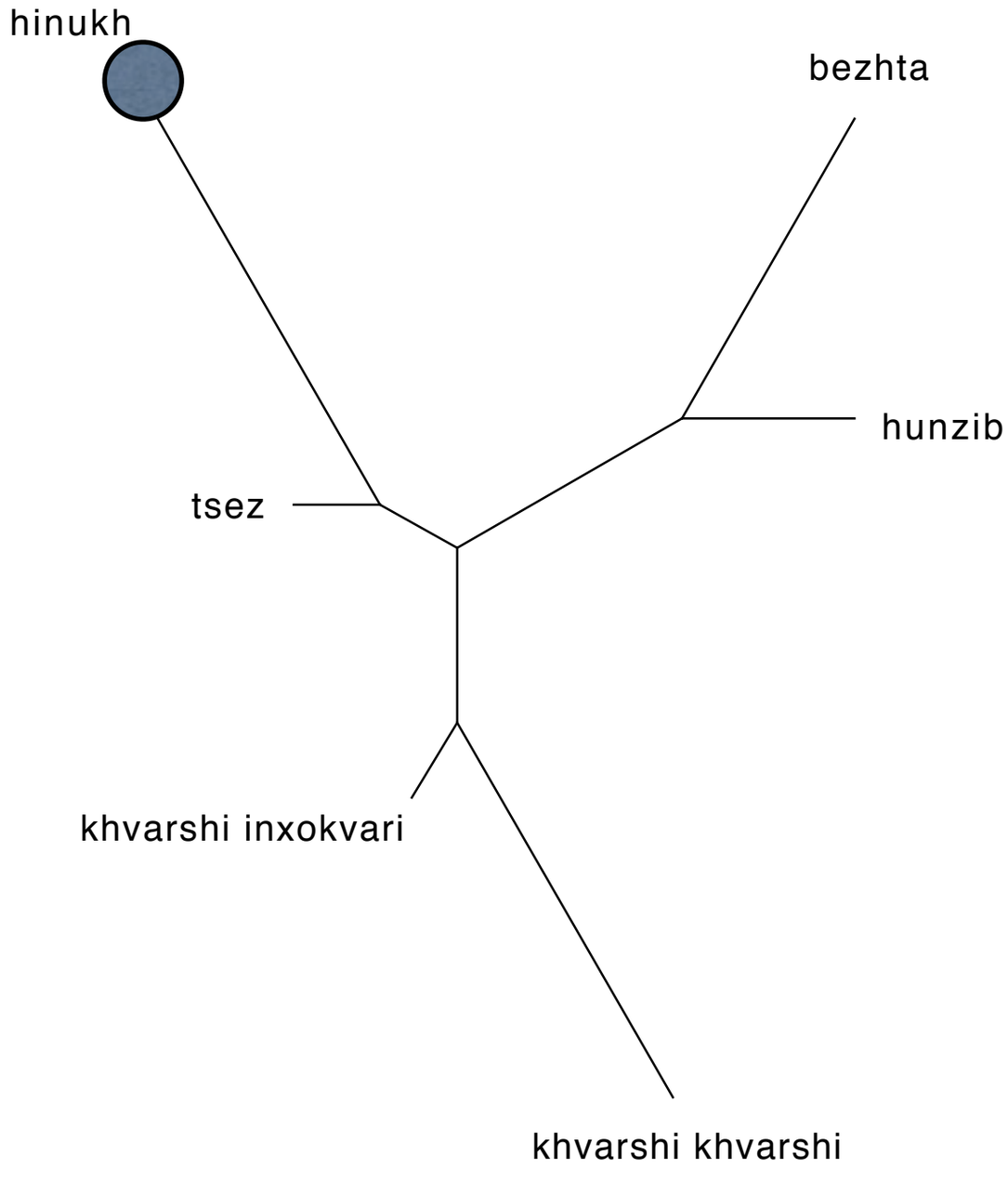

1.0
0.5
0.0
-0.5
Y
0.2
0
0.4
0.8
0.6
1
-1.5
-1.0
-0.5
0.0
0.5
X

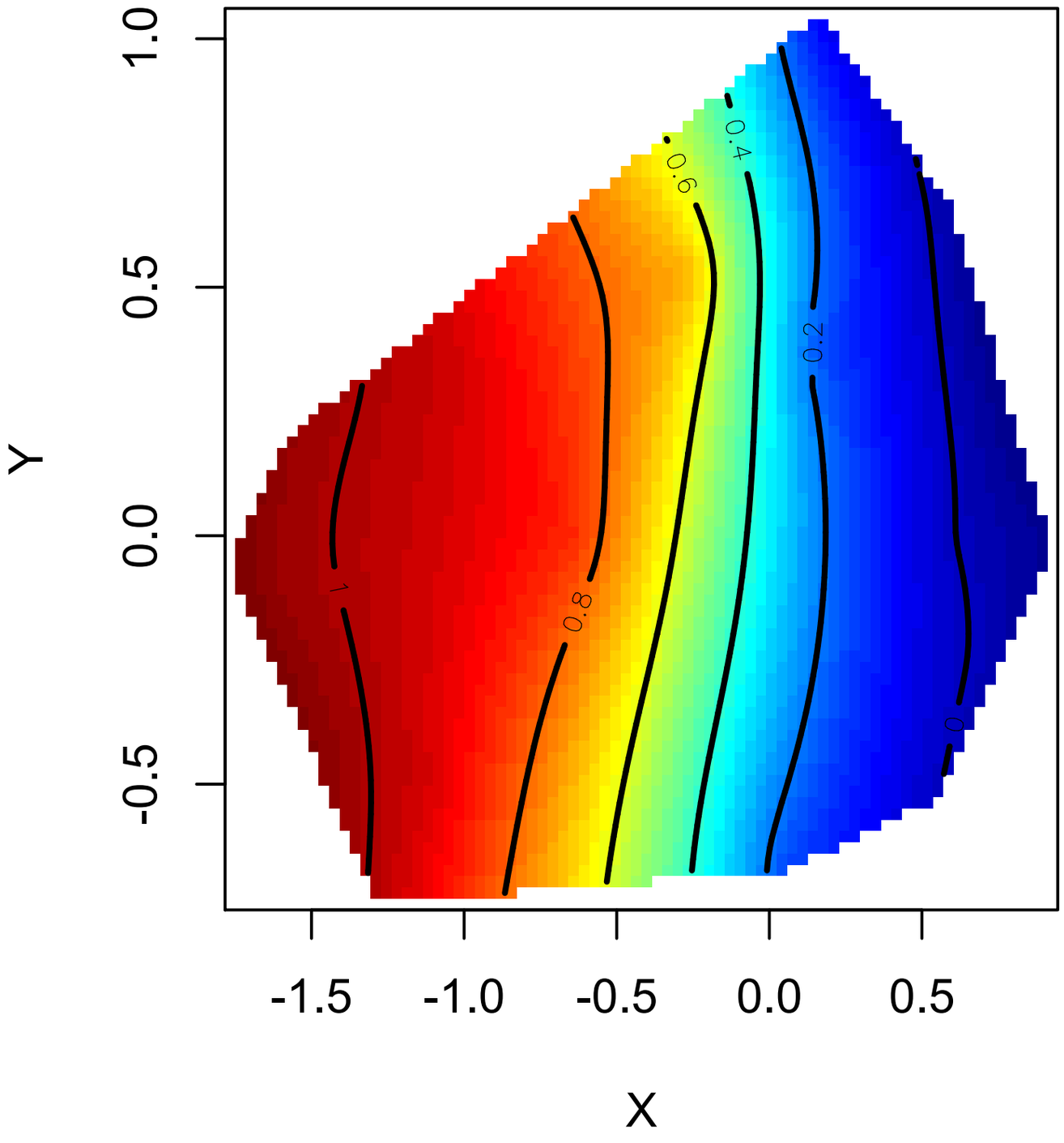

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

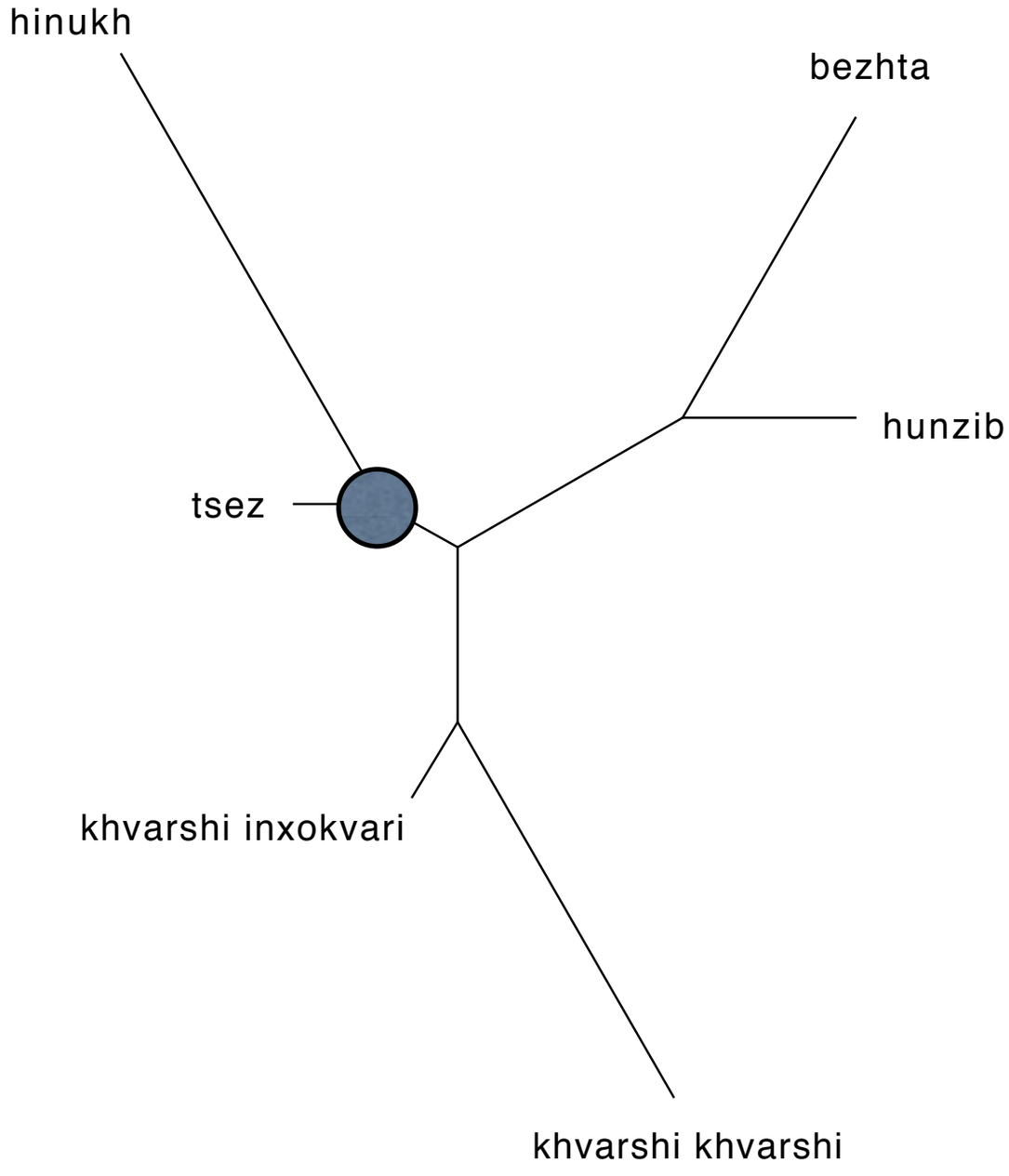

1.0
0.5
0.0
-0.5
Y
-1.5
-1.0
-0.5
0.0
0.5
X
0.4
0.6
0.2
1
0.8
0

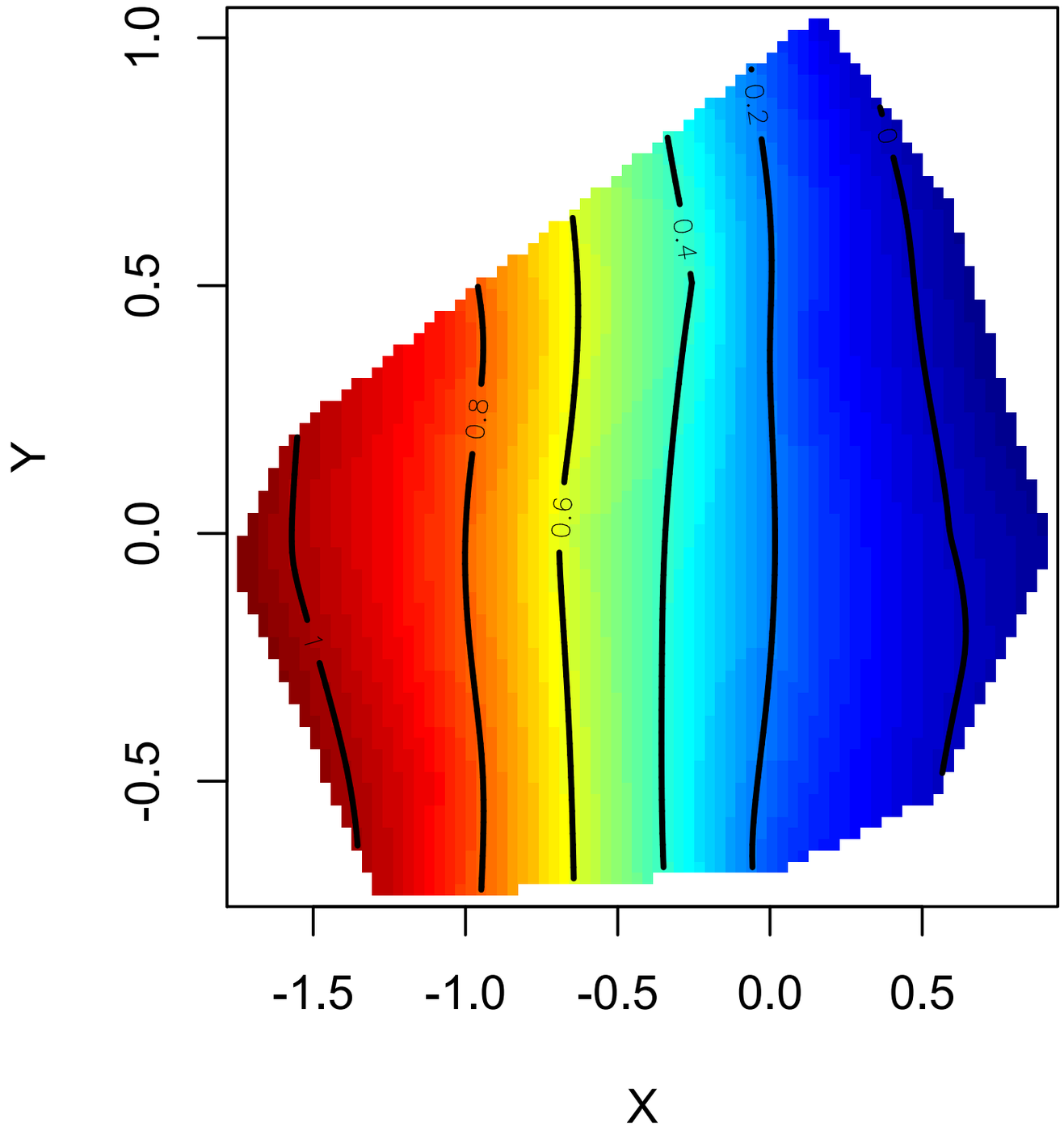

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

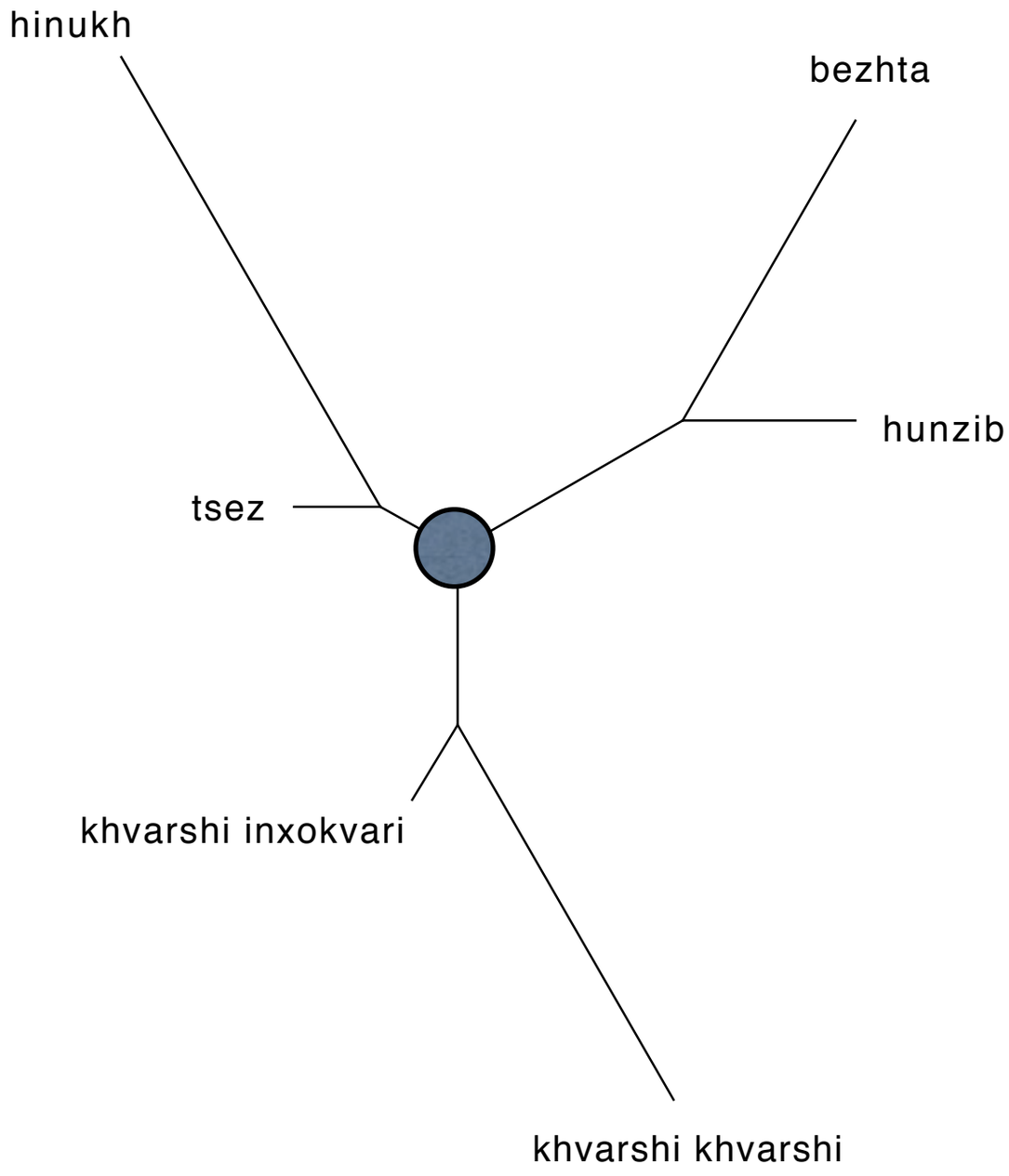

1.0
0.5
0.0
-0.5
Y
-1.5
-1.0
-0.5
0.0
0.5
X
1
0.8
0.6
0.4
0.2
0

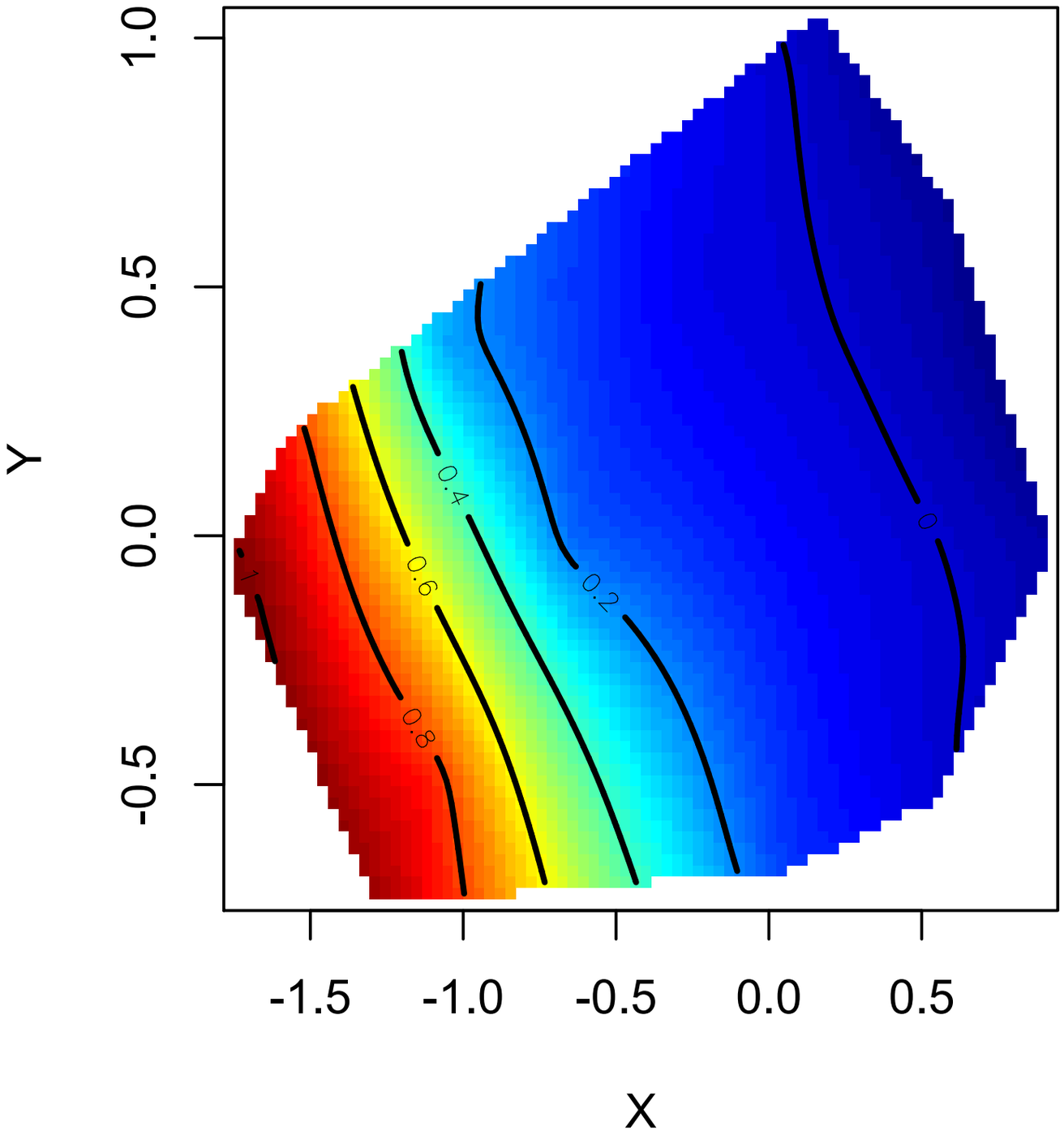

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

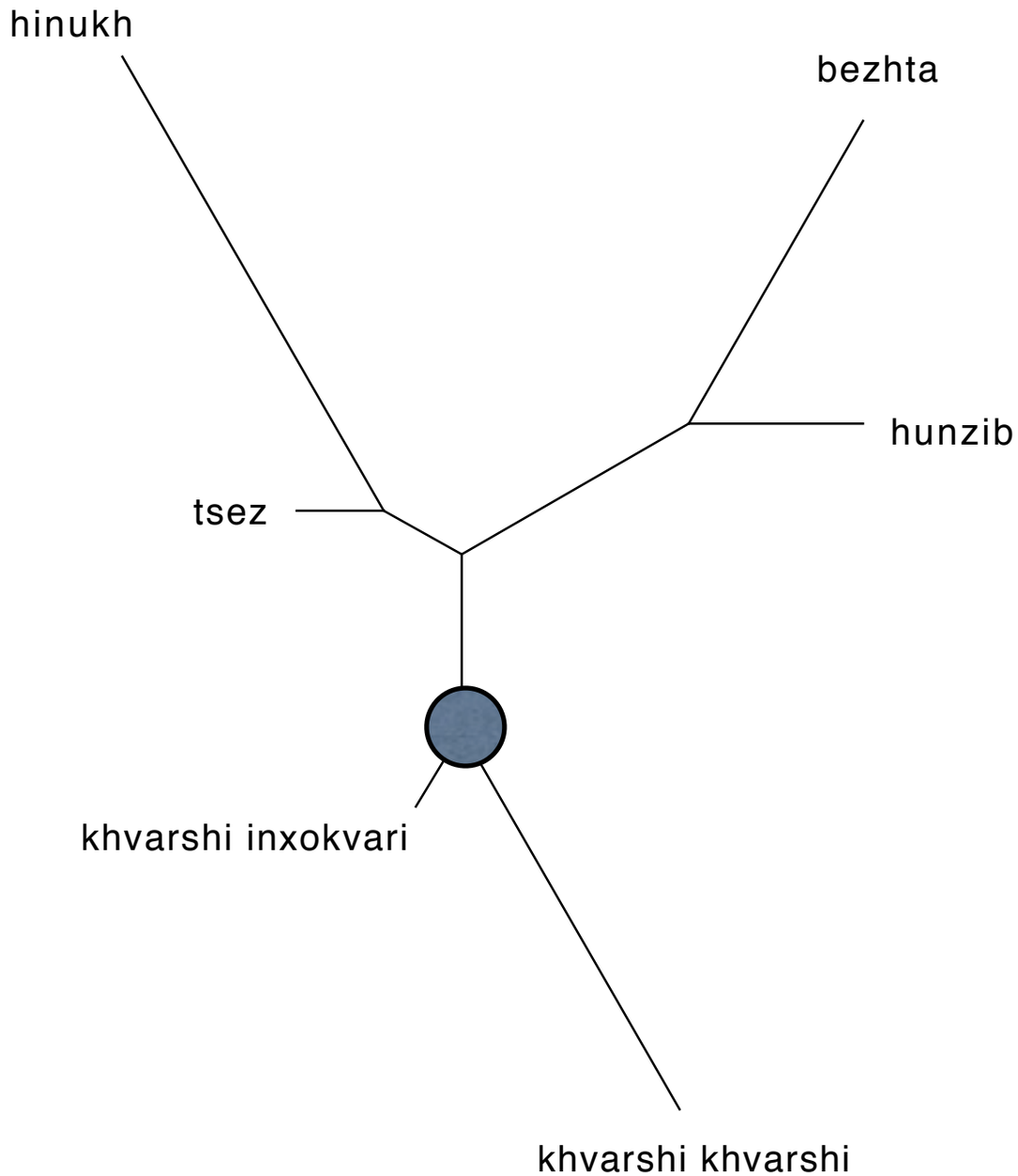

1.0
0.5
0.0
-0.5
Y
-1.5
-1.0
-0.5
0.0
0.5
X
0.4
0.6
0.2
1
0.8
0

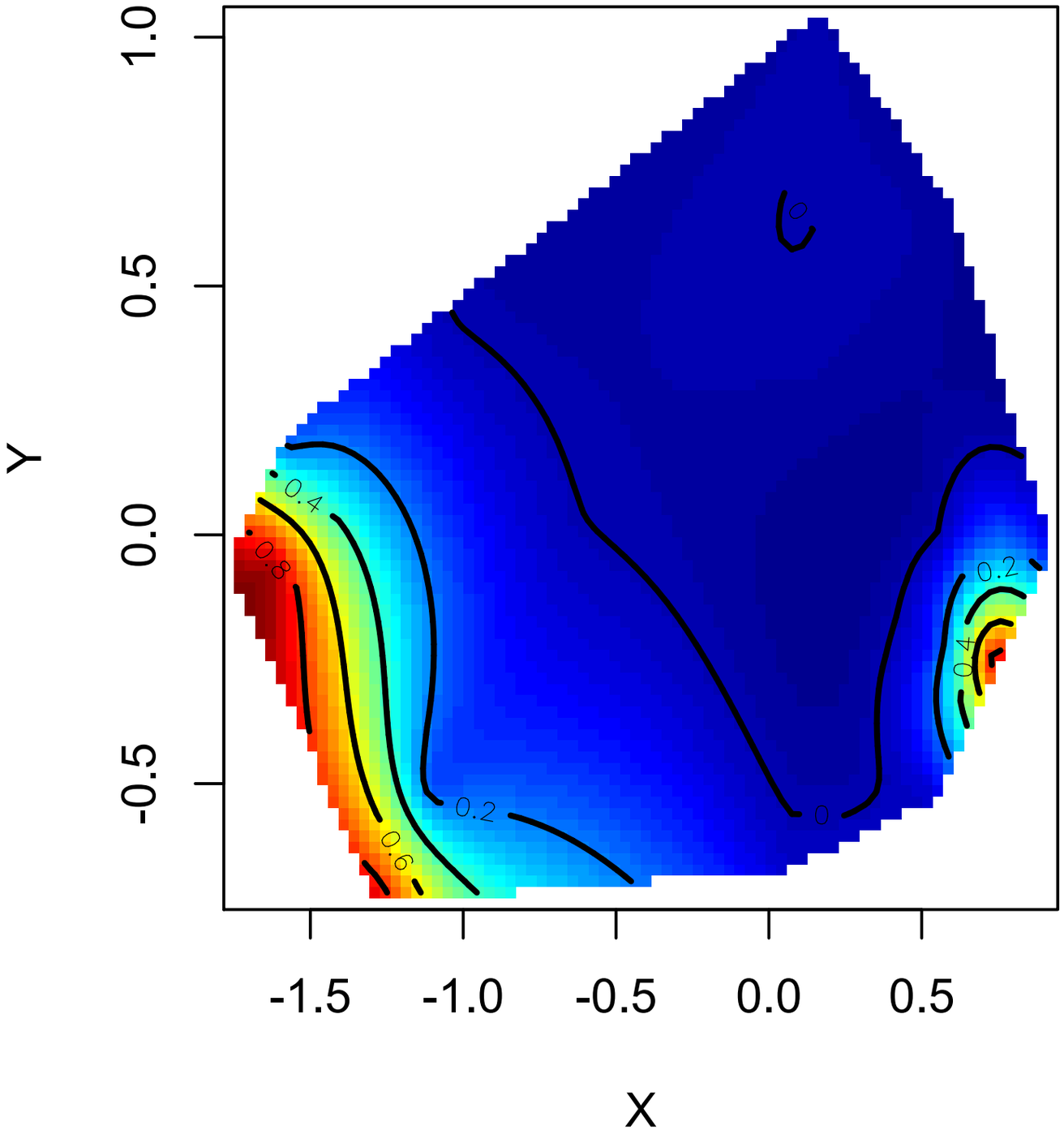

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

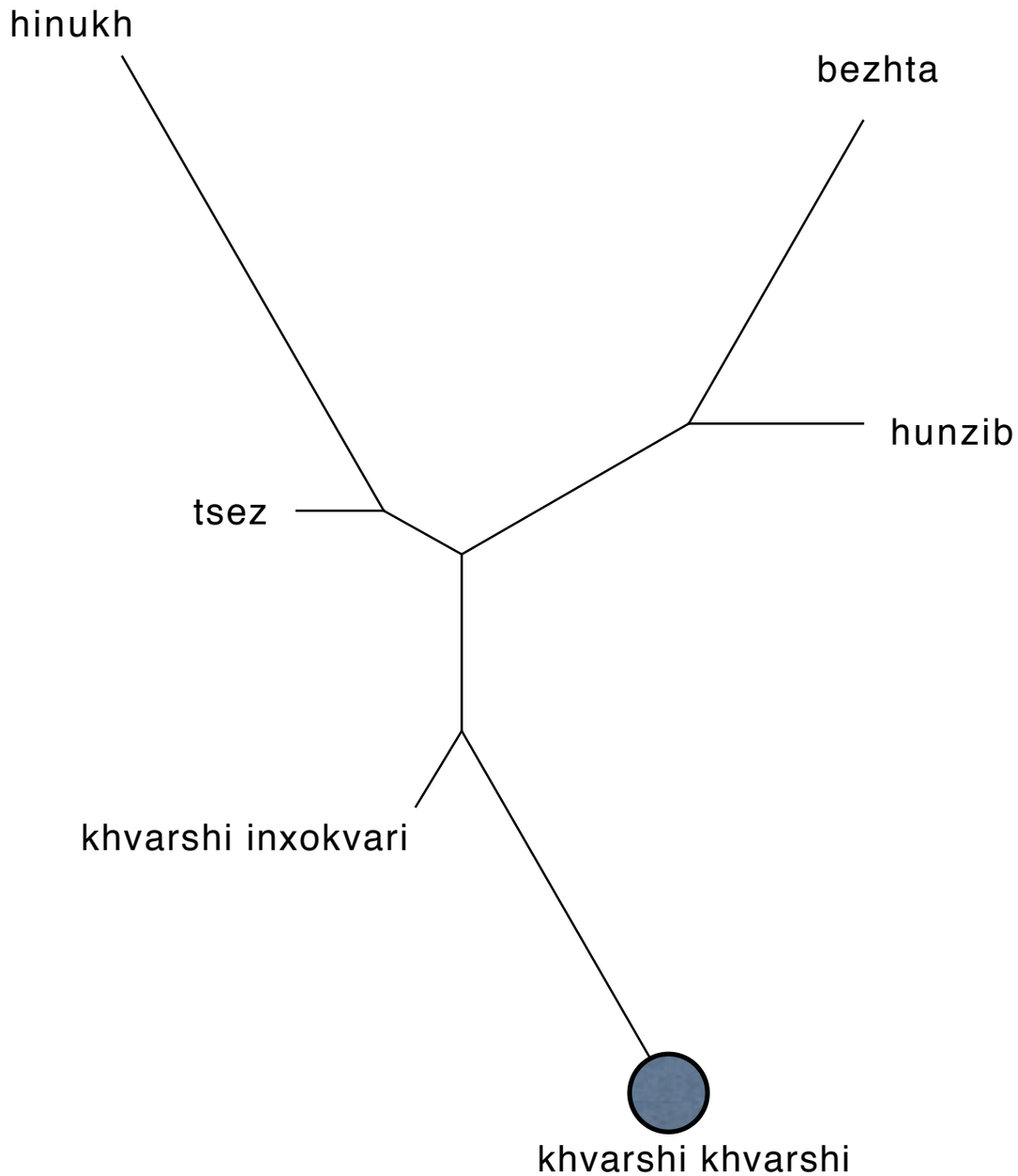

1.0
0.5
0.0
-0.5
Y
0.2
0
0.4
0.8
0.6
1
-1.5
-1.0
-0.5
0.0
0.5
X

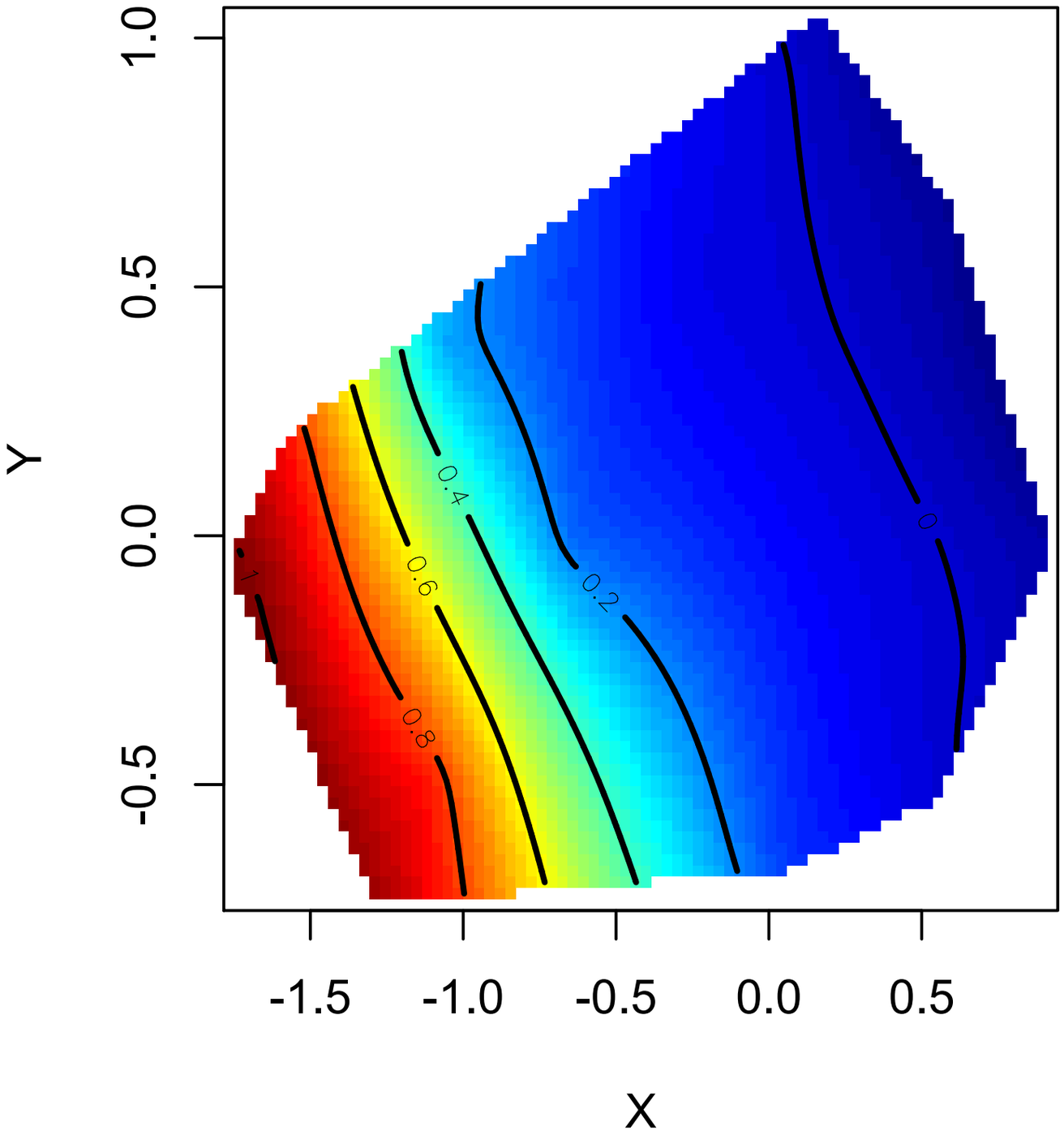

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

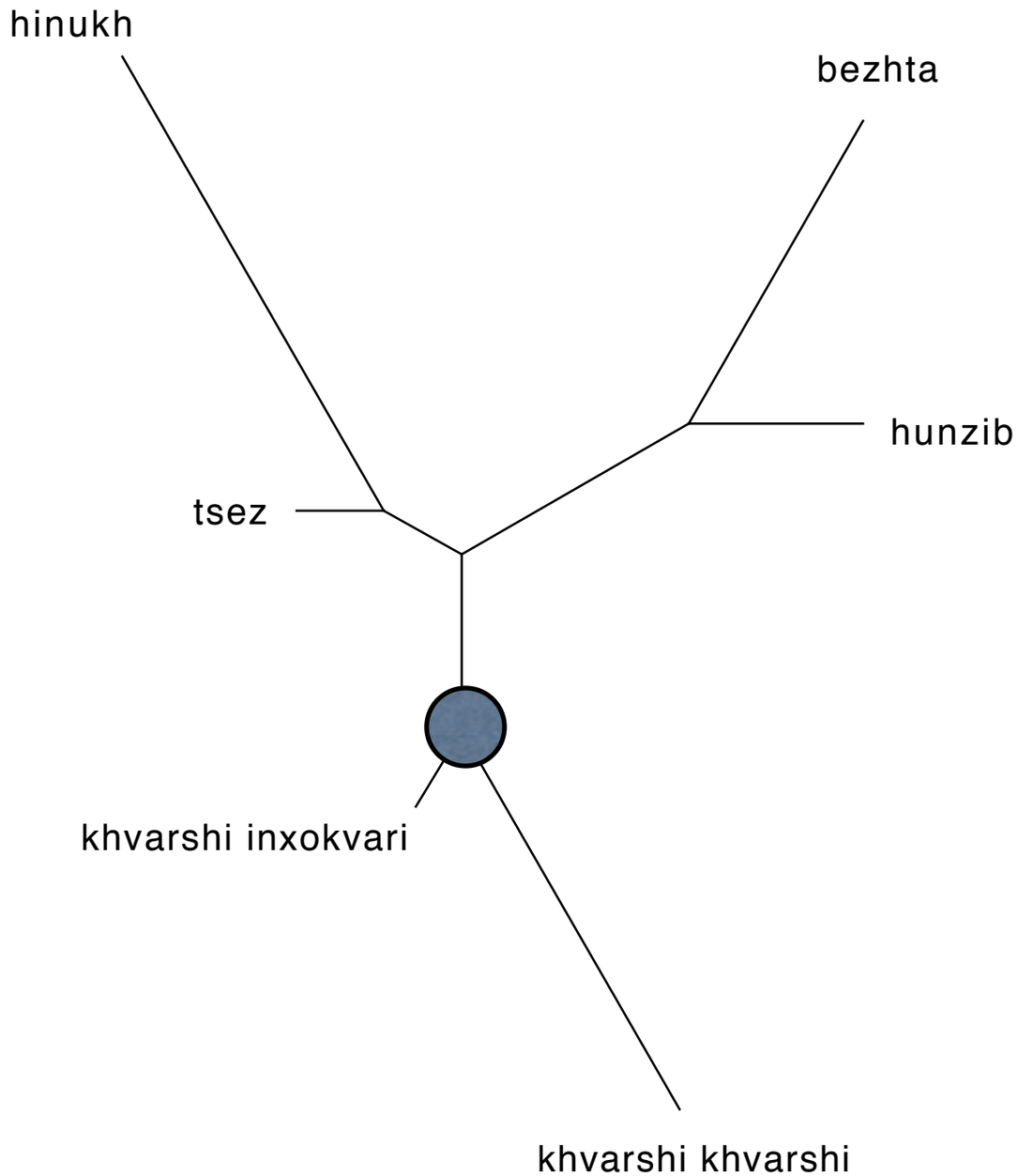

1.0
0.5
0.0
-0.5
Y
-1.5
-1.0
-0.5
0.0
0.5
X
1
0.8
0.6
0.4
0.2
0

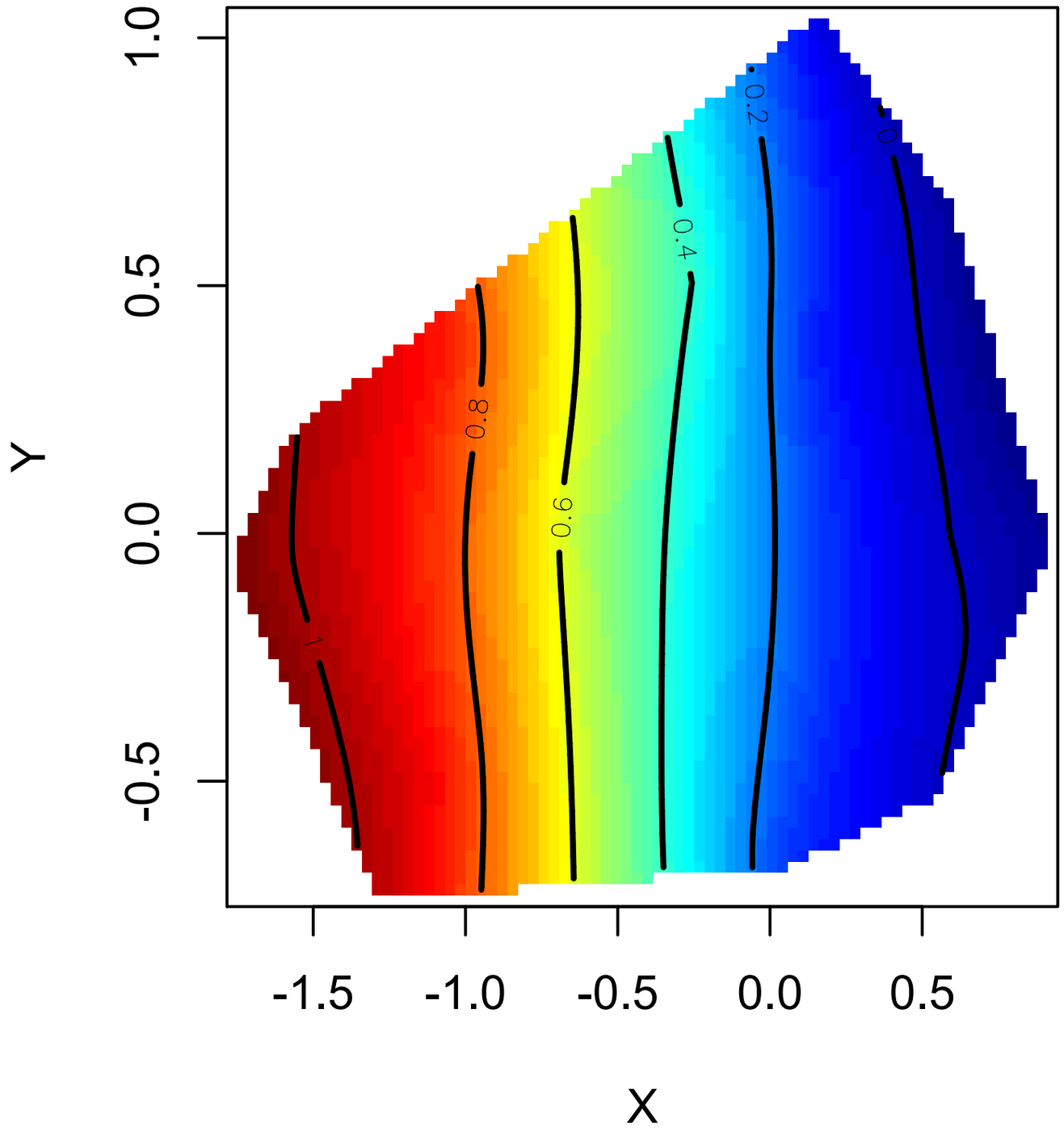

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

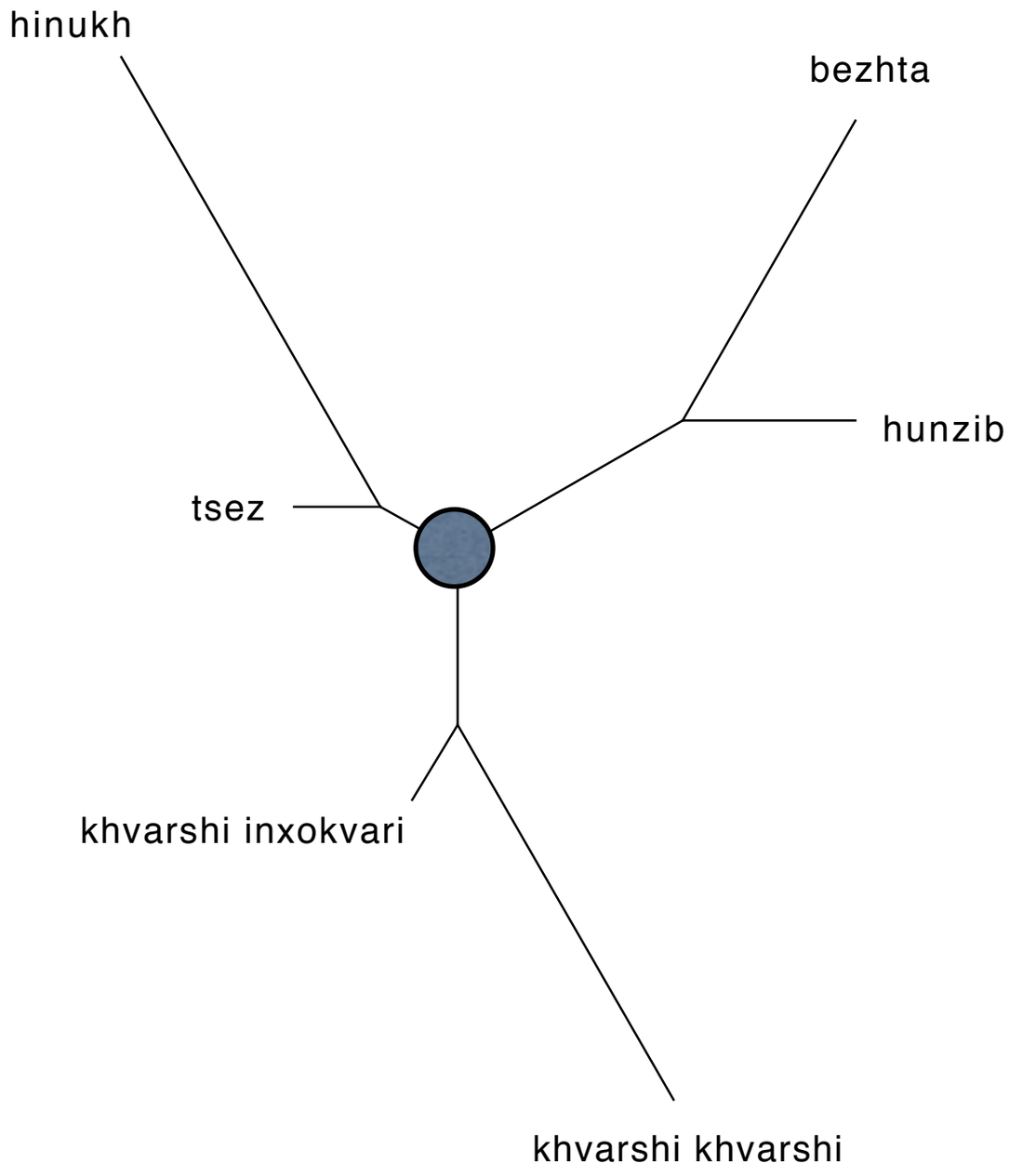

1.0
0.5
0.0
-0.5
Y
0.4
0.8
0.2
0.6
0.2
0
0.4
-1.5
-1.0
-0.5
0.0
0.5
X

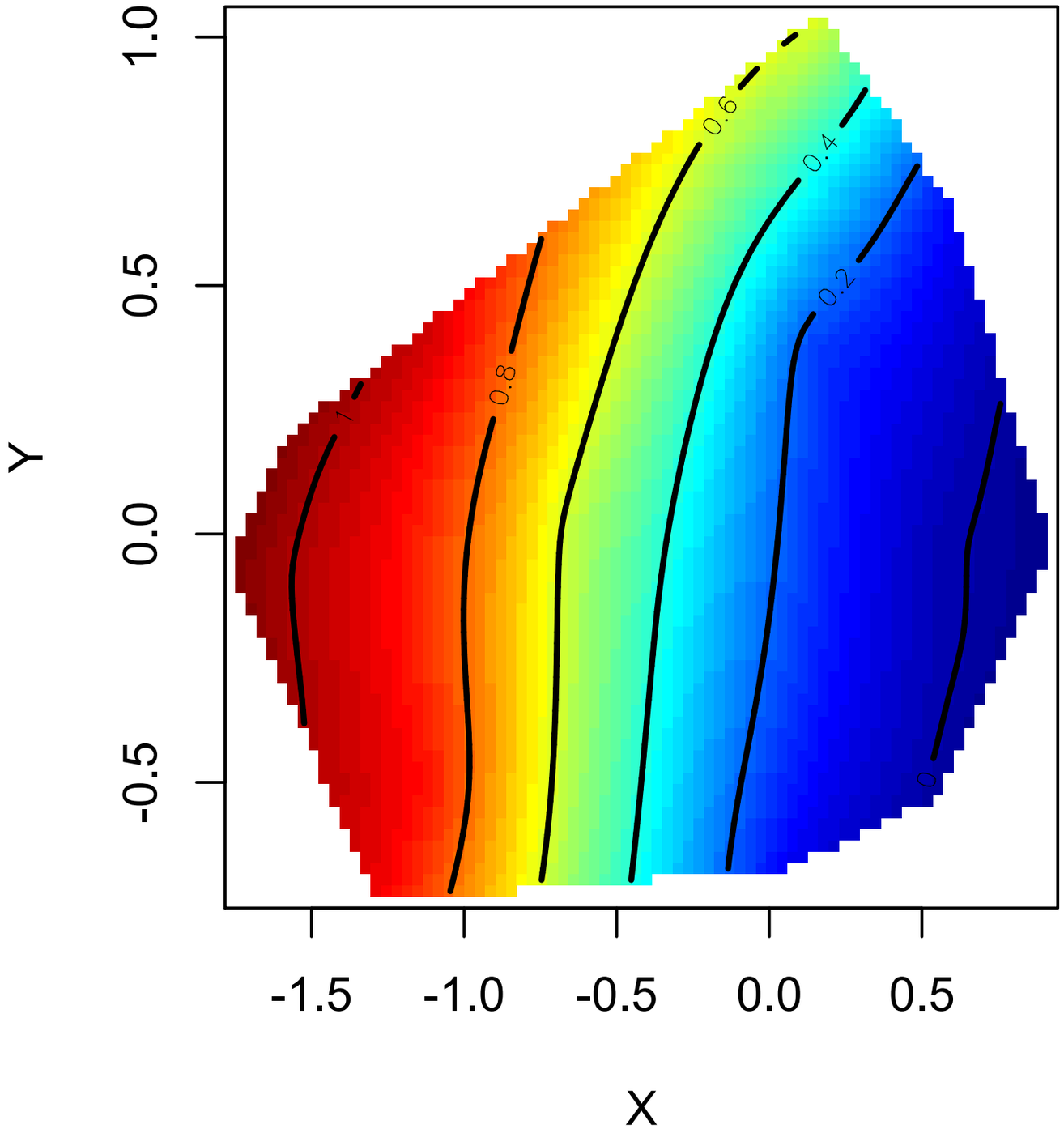

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

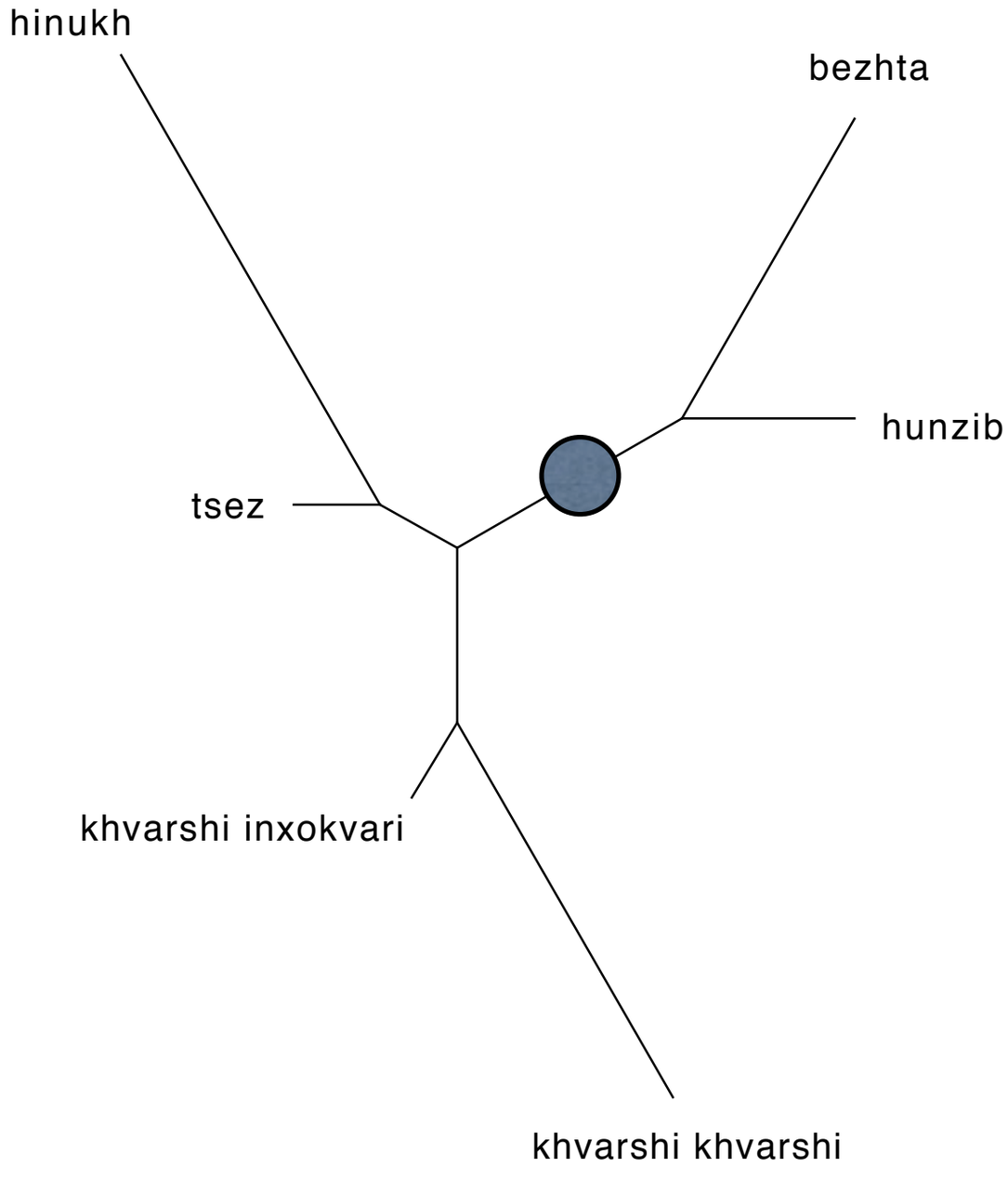

1.0
0.5
0.0
-0.5
Y
-1.5
-1.0
-0.5
0.0
0.5
X
1
0.8
0.6
0.4
0.2
0

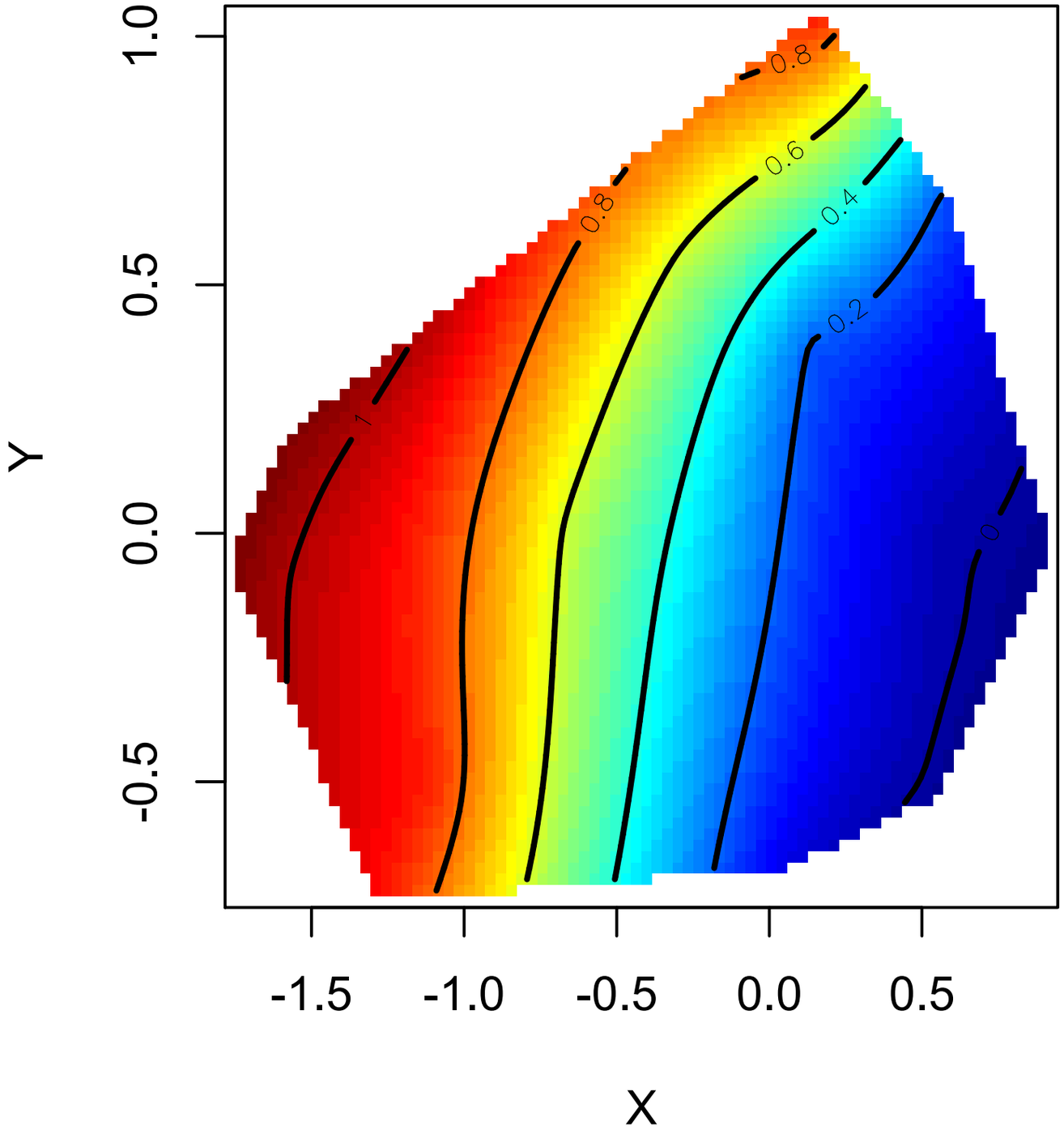

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

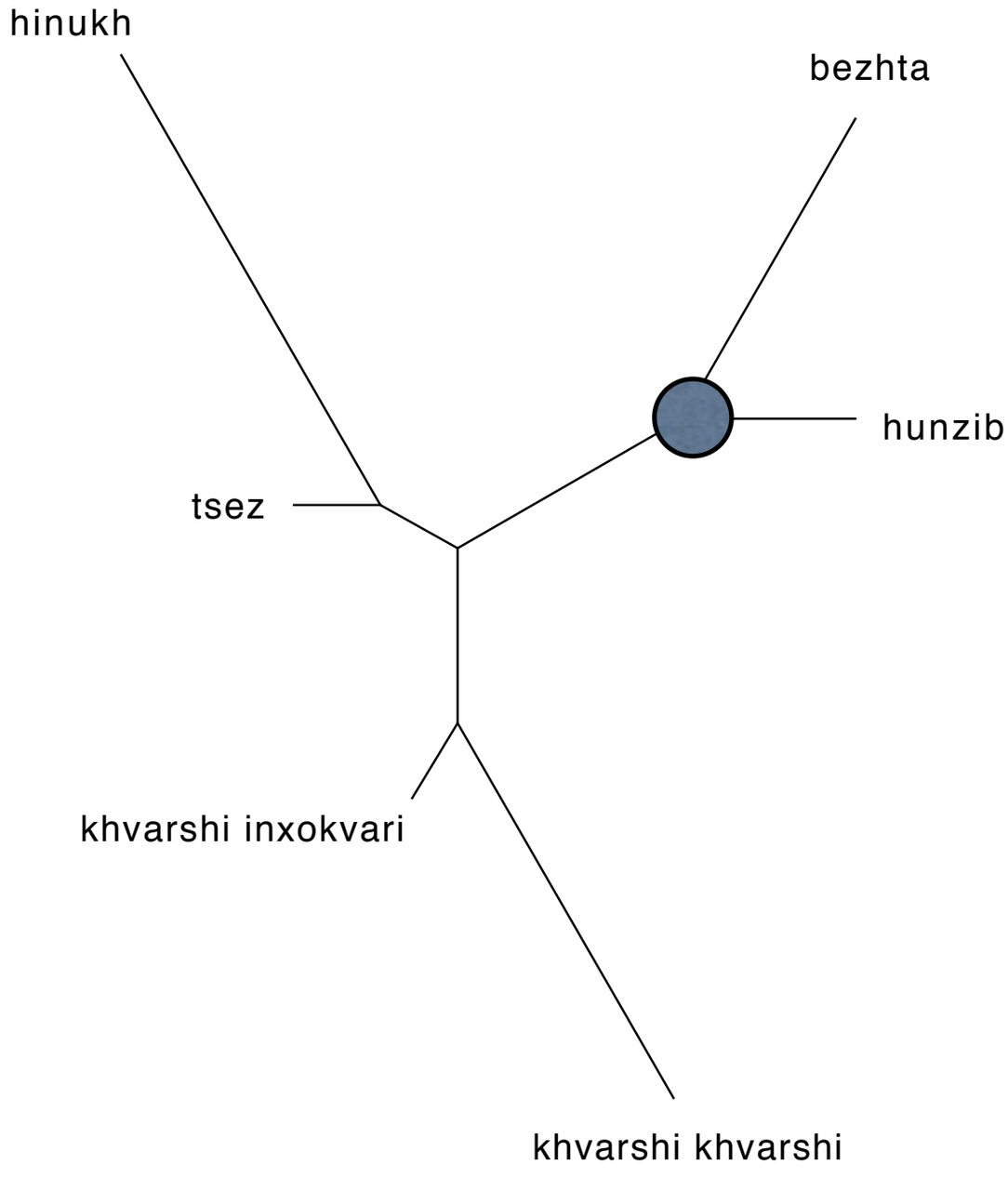

1.0
0.5
0.0
-0.5
Y
0.2
0
0.4
0.8
0.6
1
-1.5
-1.0
-0.5
0.0
0.5
X

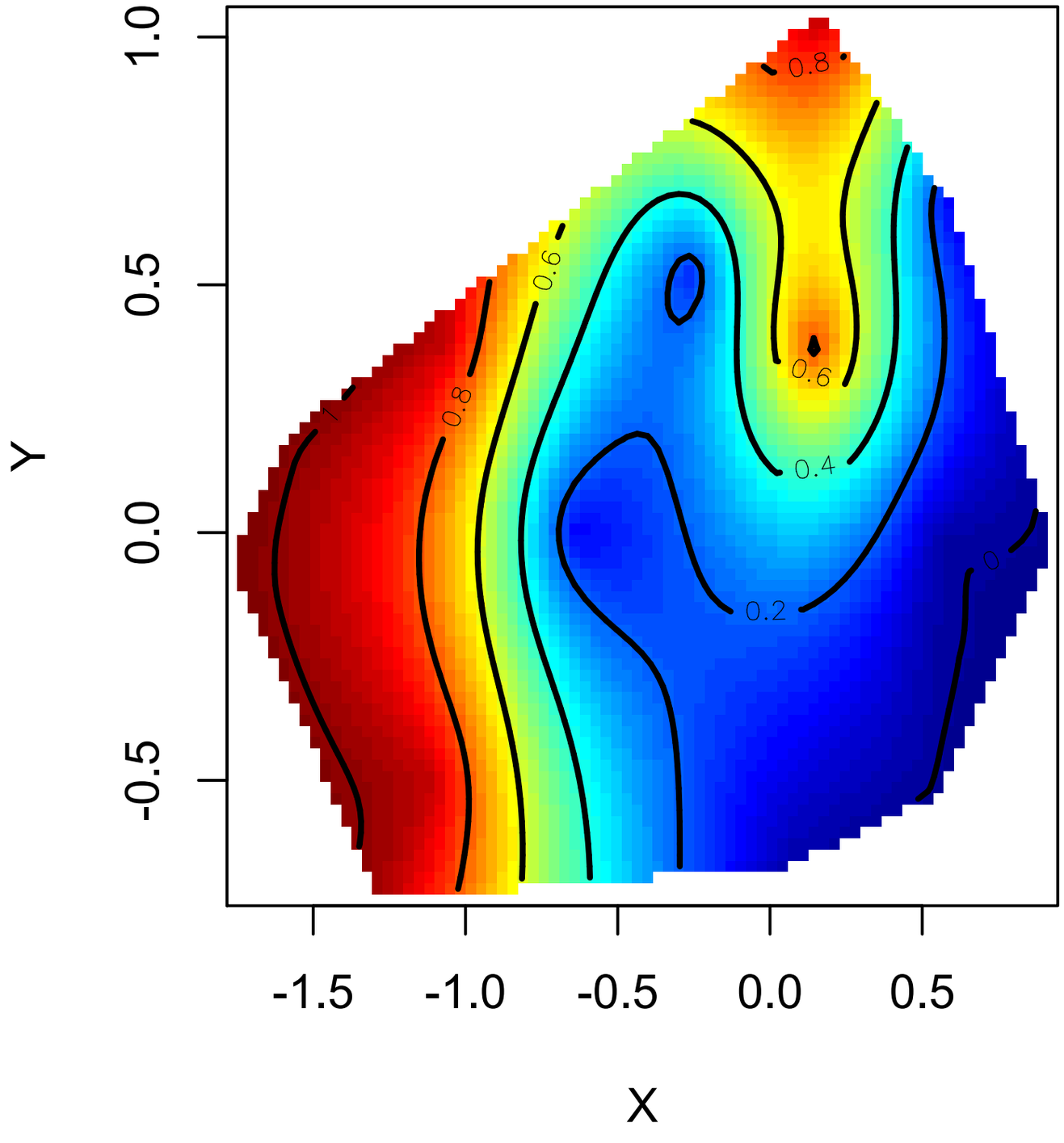

hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi

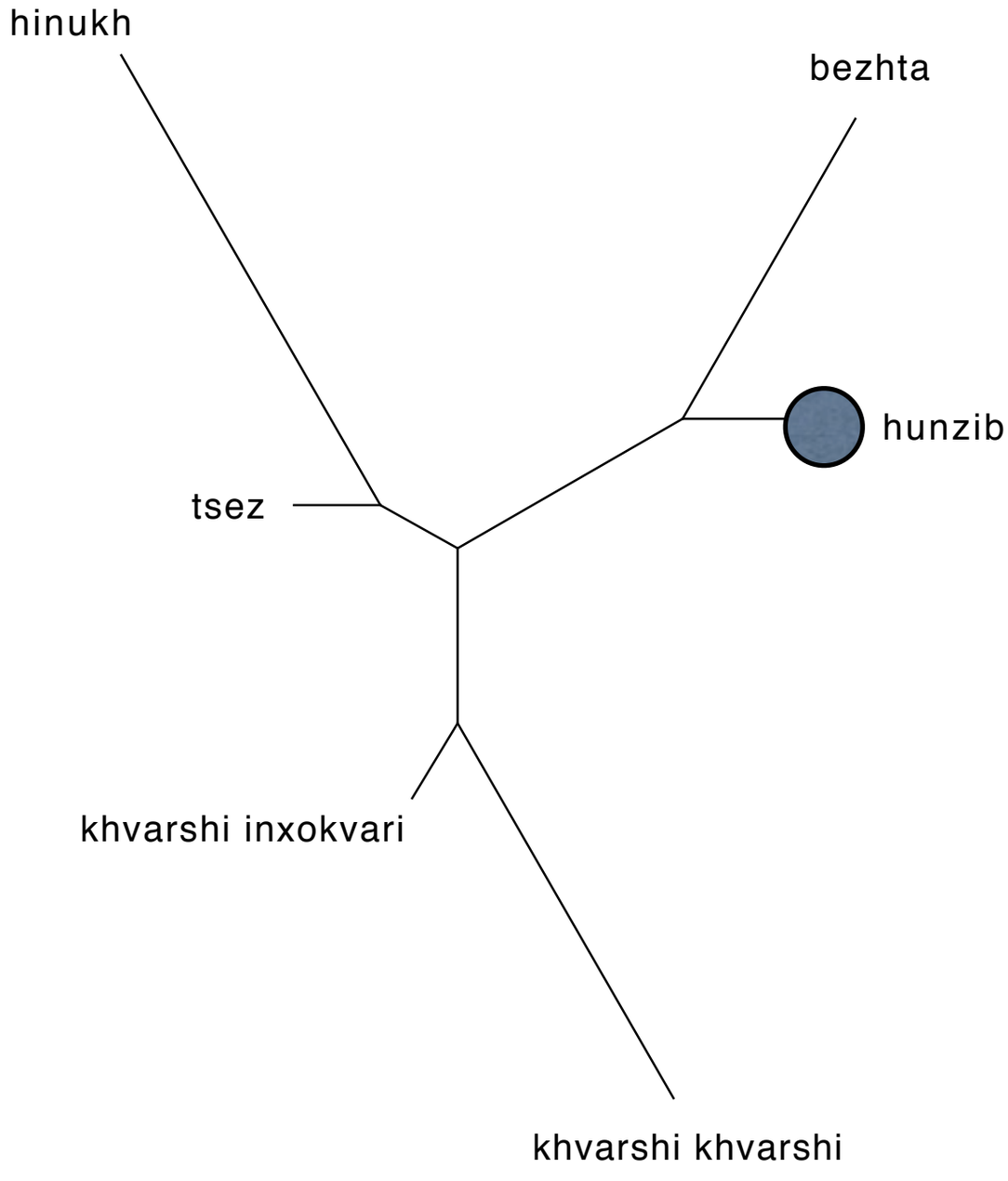

1.0
0.5
0.0
-0.5
Y
-1.5
-1.0
-0.5
0.0
0.5
X
1
0.8
0.6
0.4
0.2
0


hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi


1.0
0.5
0.0
-0.5
Y
0.8
0.6
0.4
0.2
1
0
-1.5
-1.0
-0.5
0.0
0.5
X


hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi


1.0
0.5
0.0
-0.5
Y
-1.5
-1.0
-0.5
0.0
0.5
X
1
0.8
0.6
0.8
0.6
0.4
0.2
0


hinukh
bezhta
hunzib
tsez
khvarshi inxokvari
khvarshi khvarshi