|  |  |
| --- | --- |
|  | .NET Bio Extension for Excel: Водич за кориснике |

Сажетак

Описана је употреба додатка .NET Bio Extension for Excel за Microsoft® Office Excel 2007 и Excel 2010 који омогућује једноставан и елегантан рад са геномским секвенцама, метаподацима и интервалним подацима у Excel-документу.

Неке од могућности .NET Bio Extension-а су засноване на радном оквиру .NET Bio Framework, као нпр.: скуп парсера за уобичајене формате геномских датотека, скуп секвенционих алгоритама за формирање секвенце усаглашености ДНК, скуп (софтверских) прикључака на неколико Basic Local Alignment Search Tool (BLAST) веб-сервиса за геномску идентификацију, као и интервалне геномске функције које омогућују коришћење BED датотека у Excel-у. Сам додатак .NET Bio Extension може се и програмски проширити ради искориштавања неких додатних опција NET Bio Framework-а.

Веб-адреса за преузимање .NET Bio Extension-а је http://bio.codeplex.com/.

Садржај

[Увод 3](#_Toc366349762)

[Поставка .NET Bio Extension 4](#_Toc366349763)

[Преглед корисничког прочеља 5](#_Toc366349764)

[Унос података (читање датотеке) 6](#_Toc366349765)

[Упис података (секвенци) у датотеку 8](#_Toc366349766)

[Поравнавања секвенци 9](#_Toc366349767)

[Агрегација секвенци 10](#_Toc366349768)

[Слање секвенце BLAST Веб-сервисима 10](#_Toc366349769)

[Формирање графикона расподјеле нуклеотида ДНК 13](#_Toc366349770)

[Руковање интервалним геномским подацима 14](#_Toc366349771)

[Приказ Венових дијаграма на основу интервалних геномских података 18](#_Toc366349772)

[Промјена конфигурационих опција 21](#_Toc366349773)

[Додатак A: Подржане секвенце и формати датотека 22](#_Toc366349774)

[FASTA: Секвенце података 22](#_Toc366349775)

[FASTQ: Квалитативне секвенце података 22](#_Toc366349776)

[GenBank: Формат за базе података нуклеотидних секвенци 22](#_Toc366349777)

[GFF: Generic Feature Format 23](#_Toc366349778)

[Browser Extensible Data (BED) Format 23](#_Toc366349779)

[Додатак Б: Одобравање макроа 23](#_Toc366349780)

Напомена: Овај документ је дат „онакав какав јесте”. Информације и ставови изнесени у овом документу, укључујући URL адресе и друге Интернет референце, могу се променити без претходног обавјештења. Корисник сам сноси ризик употребе.

Овај документ не пружа законска права ни за какву интелектуалну својину у било ком Microsoft-овом производу. Дозвољено је умножавати и користити овај документ за ваше интерне сврхе.

© 2011 The Outercurve Foundation.

Дистрибуирано под дозволом Creative Commons Attribution 3.0 Unported License.

Microsoft и Windows су регистровани заштићени називи фирме Microsoft Corporation. Сви остали заштићени и регистровани заштићени називи који нису наведени су власништво одговарајућих компанија или особа.

# Увод

.NET Bio Extension for Excel је намијењен биолозима који су заинтересовани или већ користе Microsoft Office Excel у својим истраживањима. Флексибилност Excel-а допушта формирање и уграђивање модула посебне намјене, као и формирање одговарајућег прочеља самог модула – картице на ленти – у овом случају картице „.NET Bio”. Додатак, сам по себи, не утиче на остале могућности Excel-a, тј. све уобичајене функције Excel-а остају на располагању, при чему је сад, конкретно, биолозима обезбијеђено још мало радног оруђа, изричито усмјереног на биоинформатику.

Програмерски је могуће додатно проширивати .NET Bio Extension користећи, као што је већ споменуто, .NET Bio Framework, а и неки други одговарајући биониформатички радни оквир. Могуће је, на примјер, програмски развити специфичне биоинформатичке апликације користећи библиотеке .NET Bio Framework-а и додати одговарајуће графичке елементе везане за апликације корисничком прочељу Biology Extension-а.

Генерално, Biology Extension омогућује:

* Коришћење датотека које садрже податке о геномским секвенцама. Подржани формати датотека су:

BED GenBank  
FASTA GFF  
FASTQ

* Поравнавање читавих или пак дјелимичних геномских секвенци. Подржани алгоритми су:

MUMmer 3.0 Pairwise-Overlap (Reference Implementation)  
Needleman-Wunsch Smith-Waterman  
NUCmer 3.0

* Формирање приказа секвенце усаглашености на основу вишеструког поравнавања геномских секвенци.
* Слање приказа секвенце усаглашености биолошким веб-сервисима на идентификацију.  
  Подржани сервиси су:

NCBI QBLAST  
EBI WU-BLAST

* Измјену и упис геномских секвенци у датотеке. Подржани формати датотека за упис су:

FASTA   
FASTQ   
GenBank

* Руковање и приказ података о геномским секвенцама:

Руковање интервалним геномским подацима у формату UCSB BED.

Извршавање операција као што су Merge (унија), Intersect (пресјек) и Subtract (разлика) на интервалним геномским подацима.

Формирање графикона са подацима о геномским секвенцама.

Приказ Венових дијаграма интервалних геномских података користећи NodeXL за Excel.

За више информација, консултовати сљедећу документацију радног оквира .NET Bio Framework:

Bio\Doc: An Overview [.NET Bio Overview.docx]  
Programming Guide [.NET Bio Programming\_Guide.docx]

Такође, погледати и Прилог A – „Подржане секвенце и формати датотека”

# Инсталација .NET Bio Extension-а

На почетку су описани предуслови, системски захтјеви и кораци за поставку (инсталацију) .NET Bio Extension-a.

#### Предуслови

Потребна су елементарна знања о:

* избору опсега ћелија у Excel-у.
* методама и номенклатури у геномици и биоинформатици.

#### Системски захтјеви

.NET Bio Extension се може поставити на сваки рачунар који може покренути Microsoft Office 2007, што је сажето у чланку на адреси  
<http://office.microsoft.com/en-us/products/HA101668651033.aspx>

На самом рачунару се мора налазити сљедећи софтвер:

* било која верзија оперативног система Windows® који може покренути Office 2007, укључујући Windows XP Service Pack (SP) 3 и све новије верзије Windows-а
* Microsoft Office Excel 2007 или Excel 2010
* NodeXL шаблон за Excel, доступан на http://www.codeplex.com/NodeXL
* Microsoft .NET Framework Version 4.0, доступан на http://www.microsoft.com/downloads/details.aspx?FamilyID=9cfb2d51-5ff4-4491-b0e5-b386f32c0992

#### Инсталација

Програм за поставку .NET Bio Extension-а је доступан на веб-адреси <http://bio.codeplex.com/>.

Након успјешне поставке .NET Bio Extension се појављује у Add and Remove Programs као „.NET Bio Extension for Excel.”

За инсталацију .NET Bio Extension for Excel

1. Затворите све Excel-дoкументе.

2. Умножите .NET Bio Extension for Excel програм за поставку на неком од директоријума (магнетног диска).

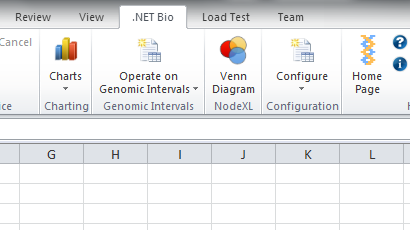
3. Отворите директоријум, а затим двапут кликните на **BioExcel.msi**, што ће покренути тзв. чаробњака (који ће вас водити кроз остатак поставке).

4. Пратите упутства чаробњака да завршите поставку Biology Extension-а.

Потврда успјешности инсталације

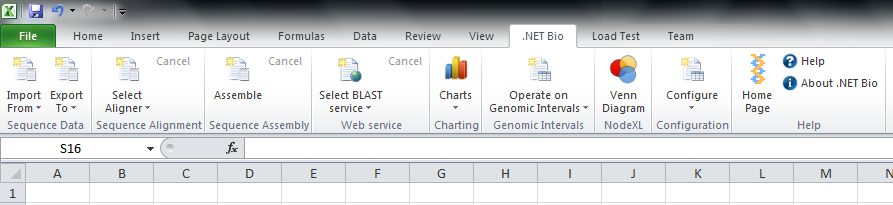
* Покрените Excel 2007 или Excel 2010.

Лента би сада требало да садржи .NET Bio картицу, као што је приказано на Слици 1.

  
Слика 1. Лента Ексела са .NET Bio картицом.

# Преглед корисничког прочеља

.NET Bio картица, као што је приказано на Слици 2, садржи главно корисничко прочеље за Biology Extension.

Слика 2. .NET Bio картица

.NET Bio картица садржи седам група наредби:

Sequence Data

Убацивање и дистрибуција геномских секвенци кроз (подржане) формате датотека (наведени у Прилогу А).

Sequence Alignment

Поравнавање цјелокупних или пак дјелимичних геномских секвенци одговарајућим (подржаним) алгоритмима.

Sequence Assembly

Састављање двије или више геномских секвенци у приказ (секвенце) усаглашености.

Web Service

Слање приказа (секвенце) усаглашености веб-сервисима ради идентификације.

Charting

Приказ графикона расподјеле сваког нуклеотида у ДНК ланцима.

Genomic Intervals

Руковање интервалним геномским подацима у формату UCSB BED помоћу наредби (операција) као што су Merge (унија), Intersect (пресјек), Subtract (разлика).

NodeXL

Приказ Венових дијаграма на основу поравнавања секвенци.

Configuration

Дефинисање сљедећих опција у Biology Extension-у:

промјена ширине колоне-оквира за податке о геномским секвенцама.

измјена шеме за бојење ДНК, РНК и протеинских молекула.

# Унос података (читање датотеке)

.NET Bio Extension подржава сљедеће типове и формате записа геномских података:

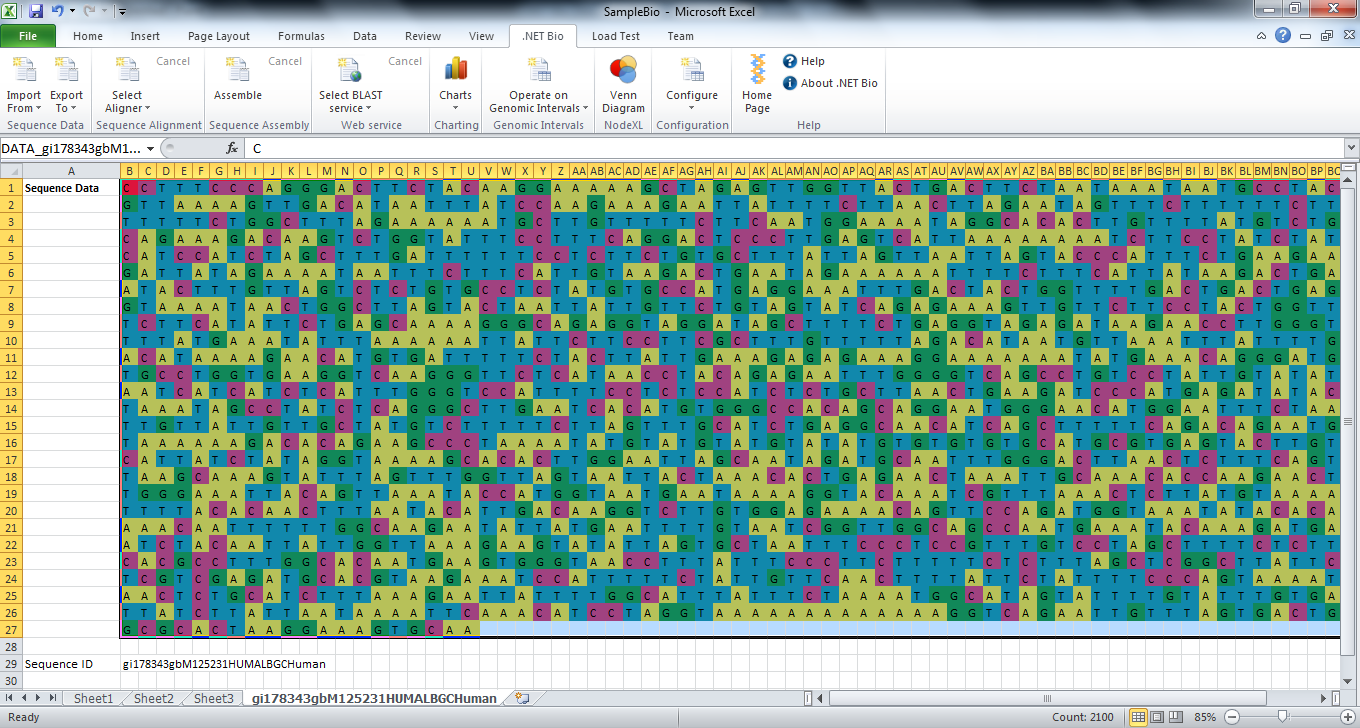
|  |  |
| --- | --- |
| Tип | Формат |
| ДНК, РНК или протеинске секвенце | FASTA, FASTQ, GenBank |
| Метаподаци о секвенцама | GFF |
| Интервални геномски подаци | BED |

Подаци се по уносу (читању датотеке) могу преуређивати, мијењати, слати BLAST веб-сервису, или пак уписивати у нове датотеке.

За унос ДНК, РНК или протеинских секвенци

1. Кликните на Import From (Убаците из..) са .NET Bio ленте.
2. Одаберите један од формата геномских података: FASTA, FASTQ, или GenBank.
3. Пронађите и одаберите жељену датотеку изабраног формата.
4. Кликните на Open.

Секвенца је унешена у нови радни лист, као на сљедећој слици.

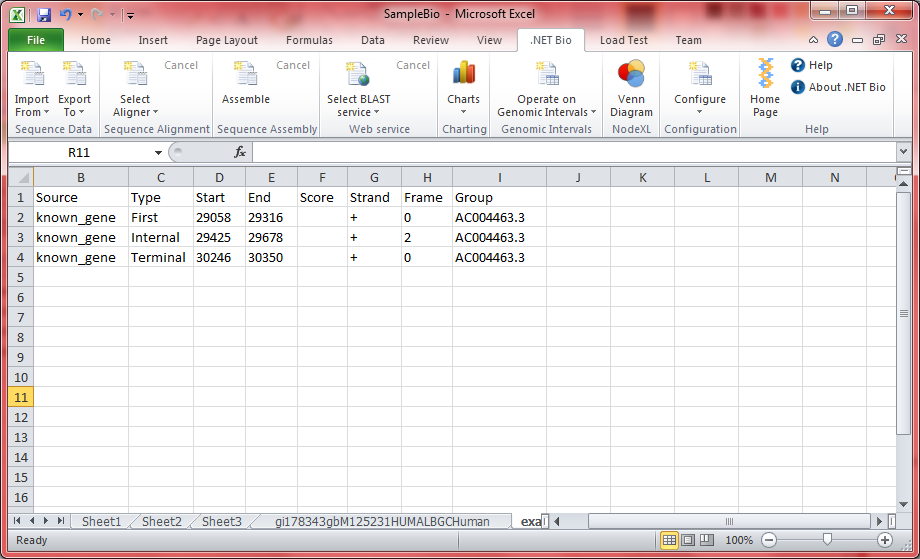


FASTA секвенца података у Excel-у

За унос GFF метаподатака о секвенцама

1. Кликните на **Import From** са .NET Bio ленте.
2. Кликните на GFF да одаберете формат GFF.
3. Пронађите и одаберите жељену датотеку изабраног формата.
4. Кликните на **Open**.

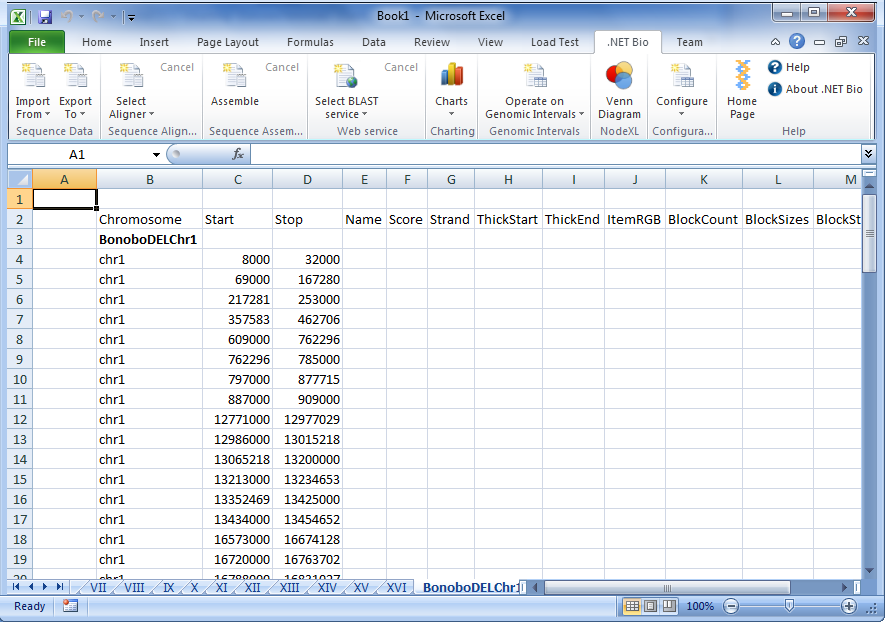
GFF метаподаци су унешени у нови радни лист, као на сљедећој слици.

  
GFF метаподаци о секвенцама у Excel-у.

За унос BED интервалних геномских података

1. Кликните на **Import From** са .NET Bio ленте.
2. Одаберите формат BED.
3. Пронађите и одаберите датотеку са интервалним подацима изабраног формата.
4. Кликните на **Open**.

Интервални подаци су унешени у нови радни лист, као на сљедећој слици.

  
**BED interval data in Excel**

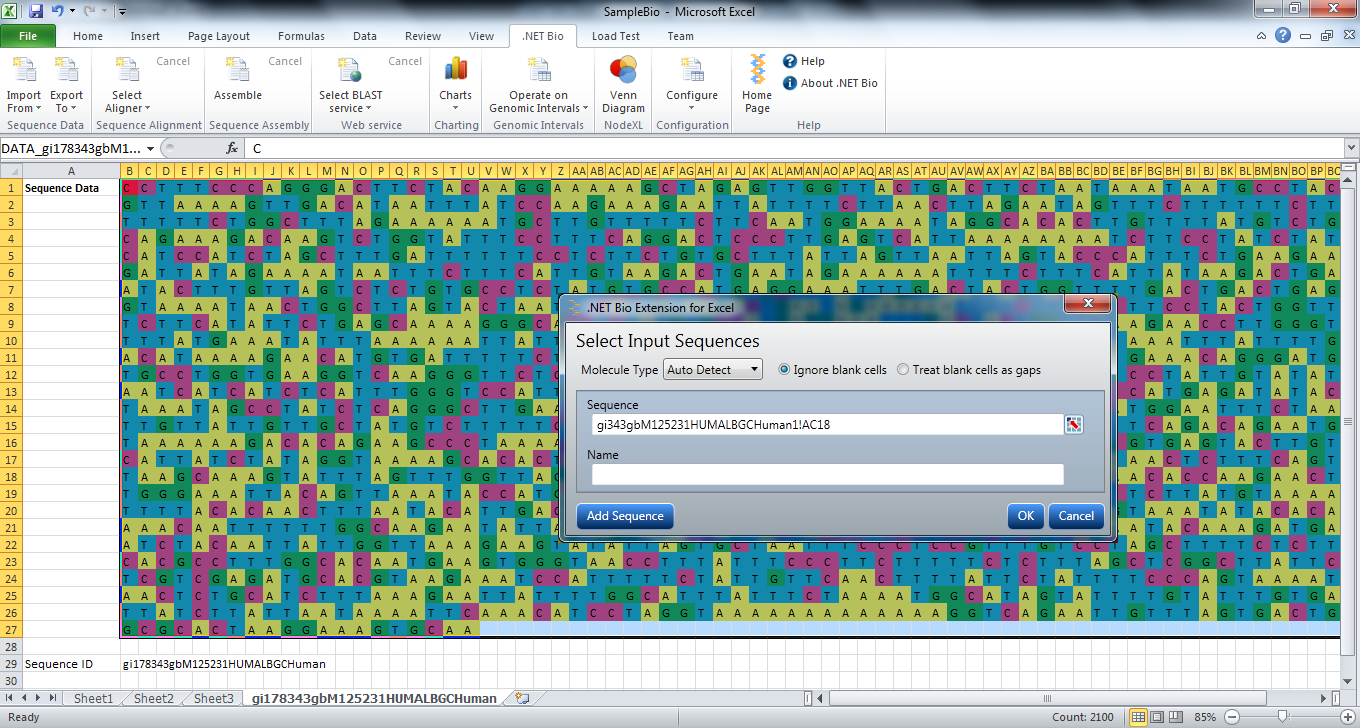
# Упис података (секвенци) у датотеку

.NET Bio Extension подржава упис ДНК, РНК или протеинских секвенци у датотеке FASTA, FASTQ и GenBank формата.

Након уноса података, могуће их је мијењати, поравнавати, слати на BLAST веб-сервис, или пак уписивати у неке друге датотеке.

За упис ДНК, РНК или протеинских секвенци у датотеку

1. Кликните на **Import From** са .NET Bio ленте.
2. Кликните на жељени формат записа секвенци: FASTA, FASTQ или Genbank.
3. Пронађите и одаберите датотеку изабраног формата.
4. Кликните на **Open**.
5. Направите жељене измјене у датотеци и кликните на Export To (Упиши у)**.**
6. Помоћу дијалошког окна Select Input Sequences (Изаберите секвенце за упис), као што је приказано на сљедећој слици, одаберите цијелу или дио секвенце и кликните на **OK**.

  
Дијалошко окно Select Input Sequences

1. Кликните на Save As на **File** картици и снимите датотеку са новим називом.

# Поравнавања секвенци

ДНК, РНК и протеинске секвенце се могу поравнавати према сљедећим алгоритмима:

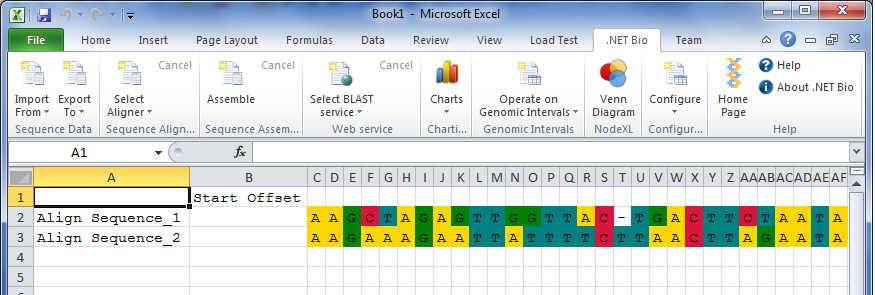
MUMmer 3.0 Pairwise-Overlap   
Needleman-Wunsch Smith-Waterman  
NUCmer 3.0

По одабиру алгоритма и двију или више секвенци, подешавате параметре за поравнавање који укључују и одабир матрице сличности. Резултат поравнавања бива приказан у новом радном листу.

За поравнавање секвенци

1. Убаците двије или више секвенци истог типа, као што је описано у поглављу „Унос података (читање датотеке)” претходно у овом документу.
2. Кликните на Select Aligner (Изаберите поравнавање) са .NET Bio ленте.
3. Одаберите двије или више секвенци у дијалошком окну Select Input Sequences, а потом кликните на **OK**.
4. Подесите параметре за поравнавање у дијалошком окну Align Inputs Parameters (Улазни параметри поравнавања) и кликните на **OK**.

Резултат поравнавања је приказан у новом радном листу, као на сљедећој слици.

Поравнате FASTA секвенце

# Агрегација секвенци

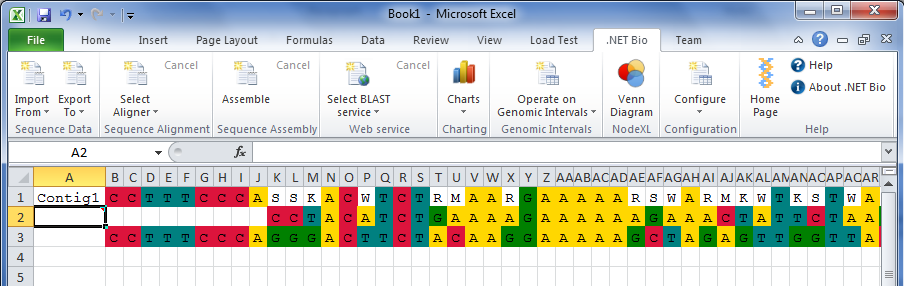
ДНК, РНК и протеинске секвенце могу бити агрегиране у тзв. приказе (секвенце) усаглашености.

Након одабира двију или више секвенци истог типа, подешавате параметре за агрегацију и бирате алгоритам поравнавања. Резулат агрегације је приказан у новом радном листу.

За агрегирање секвенци

1. Унесите двије или више секвенци истог типа како је то описано у „Упис података (секвенци) у датотеку” претходно у овом документу.
2. Кликните на **Assemble** (Агрегирај) са .NET Bio ленте.
3. Додајте унесене секвенце помоћу дијалошког окна Select Input Sequences, а затим кликните на **OK**.

4. Подестите параметре за поравнавање у дијалошком окну Align Inputs Parameters , па потом кликните на **OK**.

  
Поравнате FASTA секвенце

# Слање секвенце BLAST Веб-сервисима

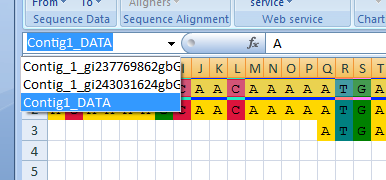
.NET Bio Extension се може искористити за слање приказа (секвенце) усаглашености, послије вишеструких поравнавања, сљедећим биолошким веб-сервисима на провјеру:

EBI WU-BLAST  
NCBI QBLAST

Након одабира цијелокупне или пак дијела секвенце и самог сервиса, подешавате параметре захтјева, који се разликују за сваки тип услуге. Резултат бива приказан у новом радном листу.

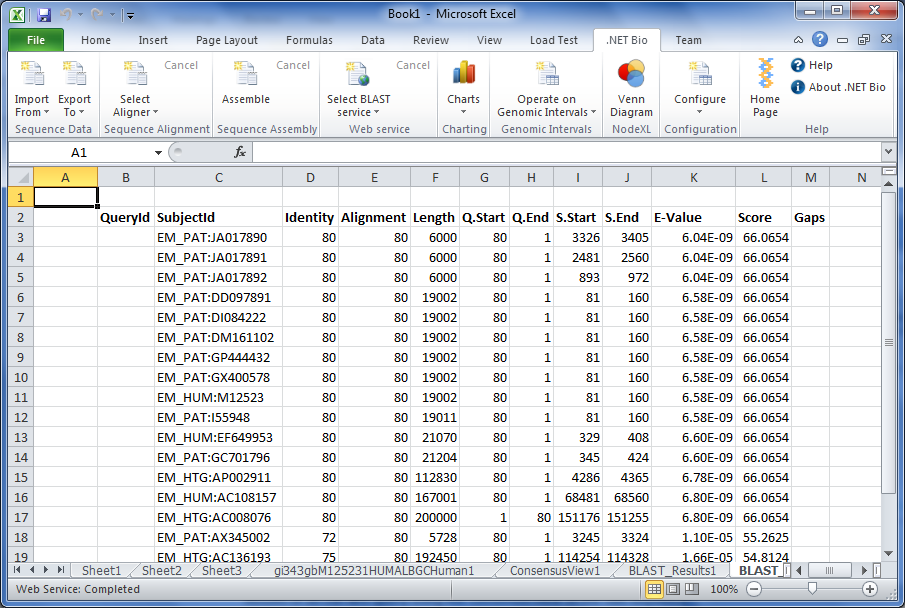
За слање цијелокупне секвенце BLAST-сервису

1. Одаберите секвенцу кликом на Contig1\_DATA у падајућем изборнику Name Box-а (Називи поља), као што је приказано на сљедећој слици.

  
Одабир consensus view двије секвенце

1. Кликните на Select BLAST Service (Изаберите BLAST-сервис) са .NET Bio ленте.
2. Кликните на одговарајући BLAST-сервис ради формирања захтјева.
3. Подесите параметре захтјева у прозору BLAST WebService (BLAST Веб-сервис) и кликните на **OK**.

Резултати су приказани у новом радном листу, као на сљедећој слици.

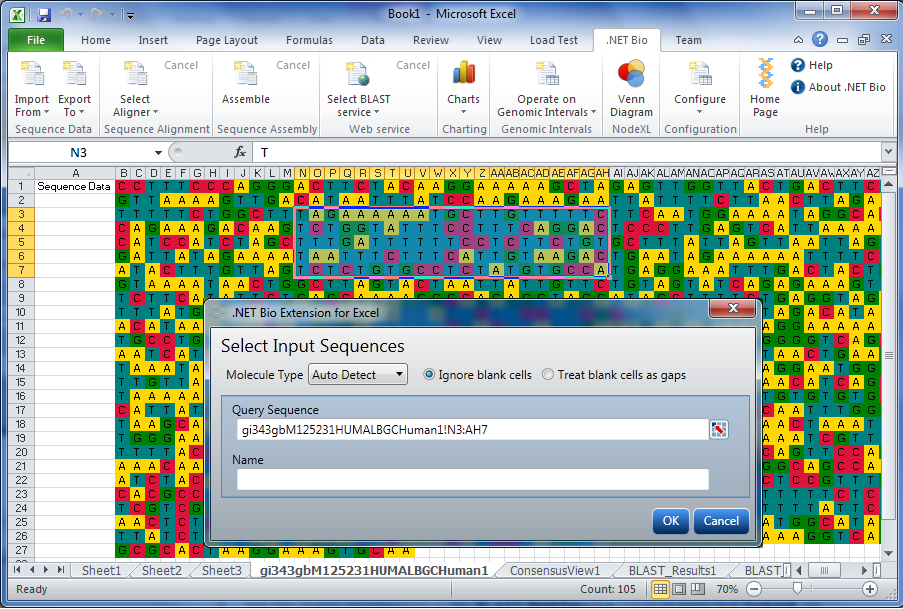
  
Резултати EBI-WU захтјева користећи приказ (секвенце) усаглашености за двије секвенце

За слање дијела секвенце BLAST-сервису

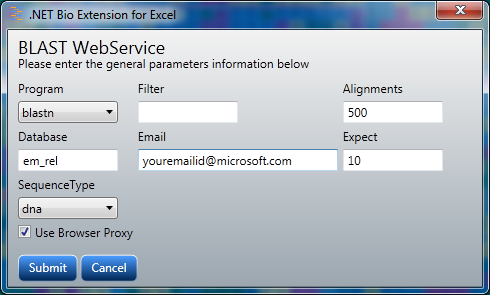
1. Одаберите, показивачем, одређене ћелије секвенце.
2. Клините на Select BLAST Service у .NET Bio ленти и одаберите BLAST-услугу.

Дијалошко окно Select Input Sequences приказује одабир у пољу Query Sequence (Секвенца захтјева), као што је приказано на сљедећој слици.

Примијетите да су посљедњи знаци ниске захтјева у ствари референце на изабране ћелије: N3:AH7.

  
Дијалошко окно Select Input Sequences

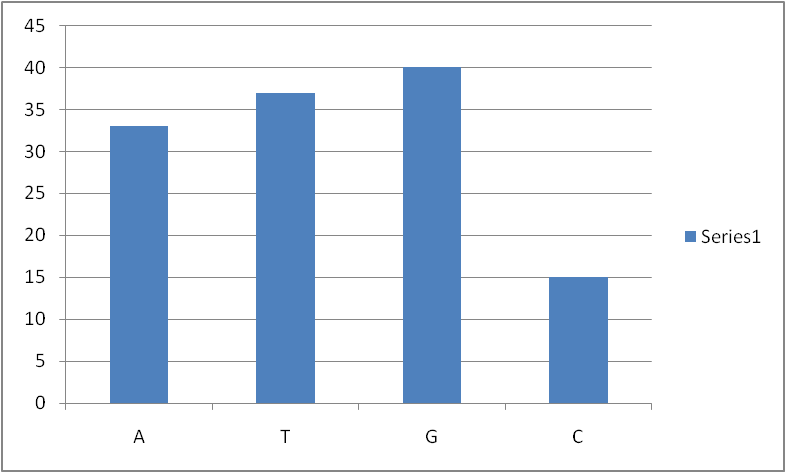
1. Подесите параметре захтјева у дијалошком окну BLAST WebService и кликните на Submit, као што је приказано на сљедећој слици.

  
Дијалошко окно BLAST WebService за EBI WU-BLAST

Ако је обрада захтјева успјешна, резултат ће бити приказан у новом радном листу.

# Формирање графикона расподјеле нуклеотида ДНК

Charting-функцијом могуће је направити графикон расподјеле нуклеотида ДНК на основу података из секвенце, као што је то приказано на Слици 3.

  
Слика 3. График расподјеле нуклеотида ДНК

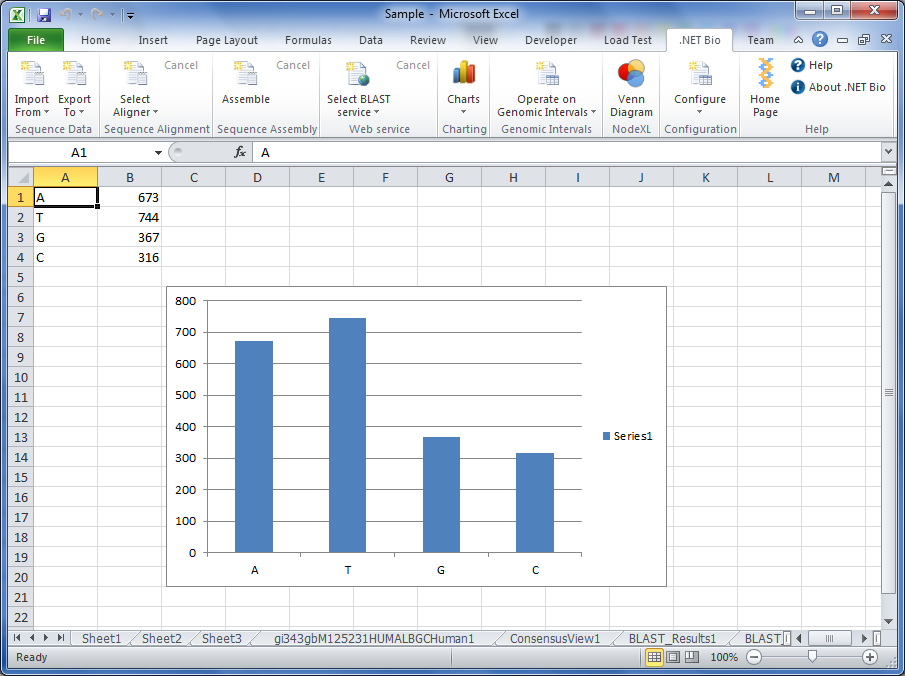
Да би могли користити Charting-функцију, морате омогућити макрое у Excel-у и додати одговарајући Excel-макро радном листу који користитe са Biology Extension-ом. Назив макроа је DisplayDNASequenceDistribution.bas и долази уз инсталацију .NET Bio Framework-а.

Важно: Слиједите процедуру из Прилога Б „Одобравање макроа” прије коришћења оруђа Charts (Графикони).

За приказ графикона ДНК секвенце

1. Отворите Excel-ову радну свеску у којој су допуштени макрои и која садржи макро **DisplayChart** (Прикажи графикон).
2. Одаберите радни лист који садржи податке о нисци.
3. Кликните на Charts иконицу са .NET Bio картице, а затим кликните на DNA Sequence Distribution Table (Табела расподјеле ДНК секвенце).

График је приказан у новом радном листу, као на сљедећој слици.

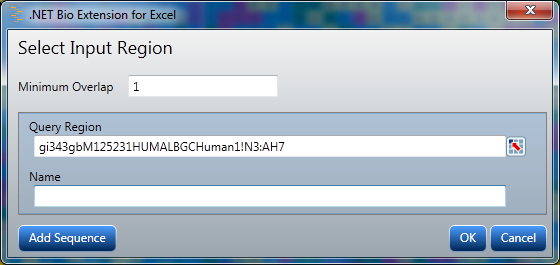
  
Нови радни лист са графиконом

# Руковање интервалним геномским подацима

Опцијом Operate on Genomic Intervals (Рад са геномским интервалима), могуће је извршити три основне (скуповне/интервалне) операције: Merge (унија)**,** Intersect (пресјек)и Subtract (разлика). Користећи датотеке у формату BED, можете дефинисати један или више захтјева користећи опсеге ћелија радног листа, а затим изабрати једну од наведене три операције. Одабрани опсези садрже једну или више базних парова координата хромозома. По обављеној операцији, резултат је уписан у нови радни лист.

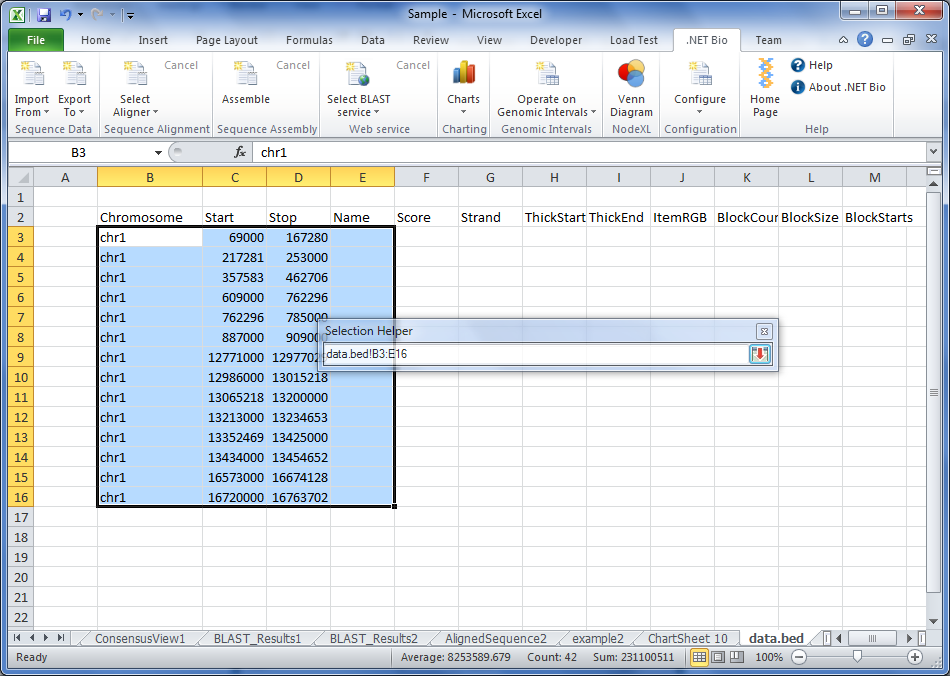
Унија преклапајућих интервала захтјева

1. Кликните **Import From** и кликните на BED.
2. Одаберите једну или више датотека и кликните на Open.
3. Одаберите радни лист и кликните на Operate on Genomic Intervals.
4. Кликните на **Мерге**. Приказано је дијалошко окно Select Input Sequences Ranges као на сљедећој слици.

  
 Прозор Select Input Sequence Ranges

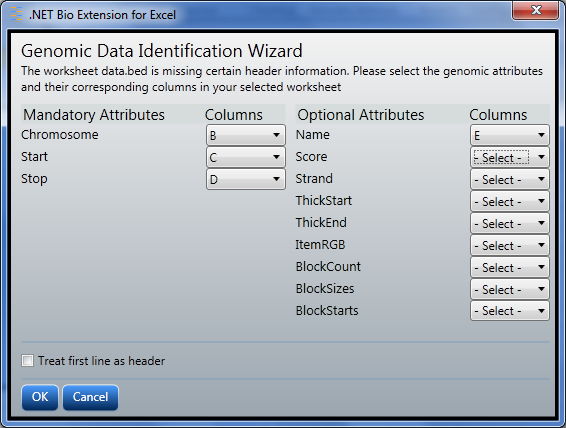
1. Кликните на иконицу за избор, која се налази уз десну страну поља Reference Sequence (Референтна секвенца) и изаберите опсег базних парова.

Примјетите да су у примјеру на сљедећој слици одабране четири колоне.

  
Избор опсега базних парова

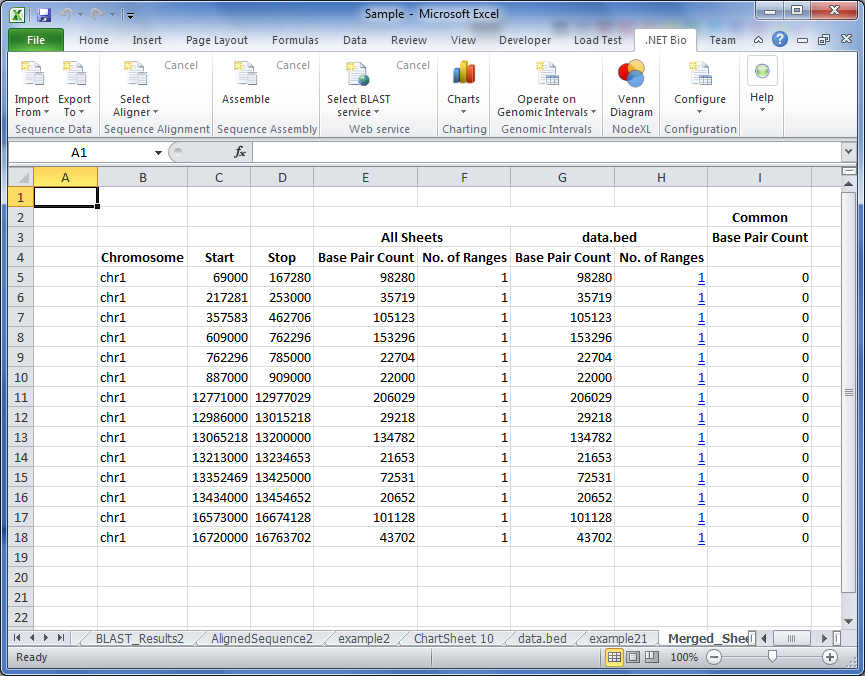
1. Клините на иконицу за избор у прозору Selection Helper (Помоћ око избора) или притисните Enter да бисте се вратили на прозор Select Input Sequences Ranges.
2. Унесите назив захтјева и кликните на OK.

Genomic Data Identification Wizard (Чаробњак за идентификацију геномских података) се приказује, као на сљедећој слици.

  
Чаробњак за идентификацију геномских података

1. Помоћу падајућих изборника подесите називе колона за четири изабране колоне — B, C, D, и E — потом кликните на OK.

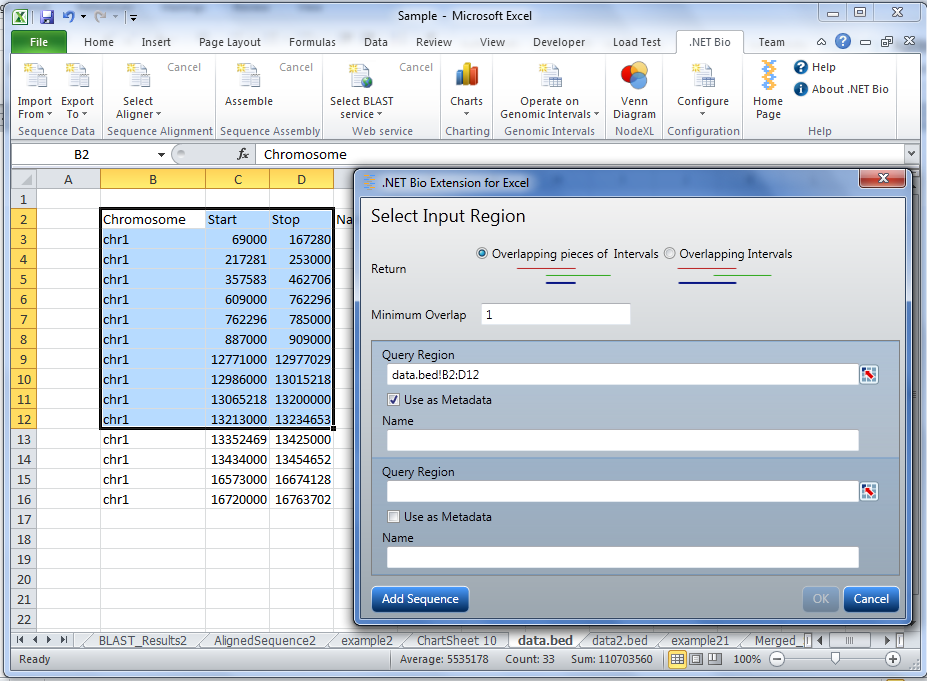
Речултати су приказани у новом радном листу под називом Merged\_Sheet1, као што је приказано на сљедећој слици.

  
Резултат уније

1. Кликните на вриједности (хипервезе) у колони No. of Ranges да бисте видјели који су опсези из првобитног радног листа обухваћени унијом.

Пресјек интервала два захтјева

1. Кликните на Import From и одаберите BED.
2. Означите једну или више датотека и кликните на Open**.**
3. Означите радни лист и кликните на Operate on Genomic Intervals.
4. Кликните на Intersect и помоћу прозора Select Input Sequences Ranges за избор два опсега базних парова: Query Region (Захтјев опсега)**,** као што је приказано на сљедећој слици.

Секвенца захтјева и референтна секвенца

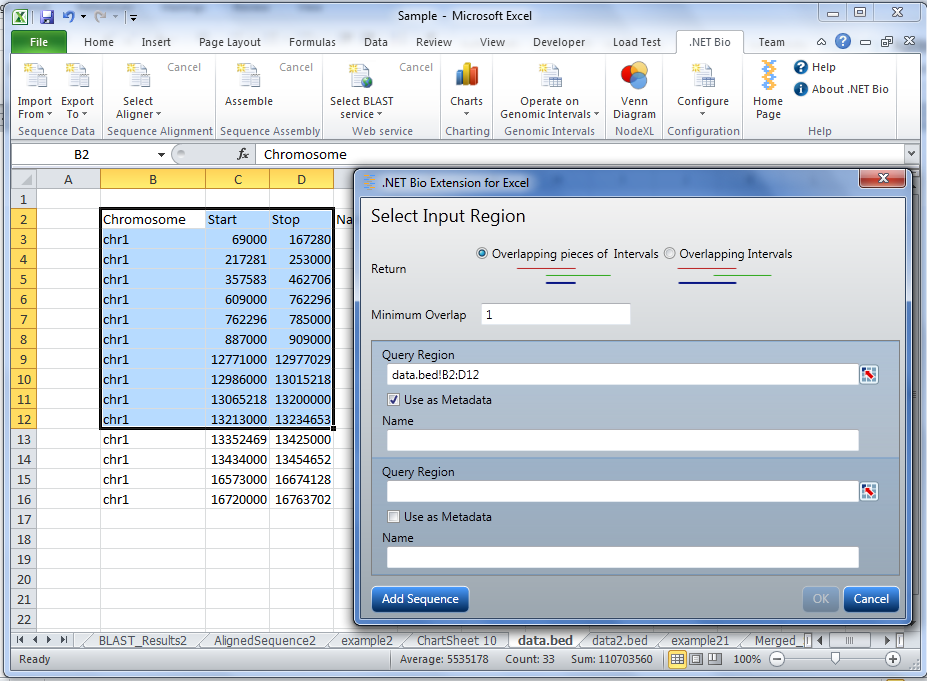
У овом примјеру се користе два радна листа, за сваку секвенцу по један.

Савјет: Додавањем заглавља избору (суб)секвенце, можете заобићи чаробњака за идентификацију геномских података. .NET Bio Extension аутоматски попуњава вриједности колоне.

1. Кликните на **OK**. Резултати ће бити приказани у новом радном листу под именом Intersect\_Sheet1.

Разлика интервала два захтјева

1. Кликните на Import From и означите BED**.**
2. Означите једану или више датотека и кликните на Open.
3. Означите радни лист и кликните на Operate on Genomic Intervals.
4. Кликните на Intersect и користите прозор Select Input Sequences Ranges за избор два опсега базних парова: Query Region, као што је приказано на сљедећој слици.

 Секвенца захтјева и референтна секвенца

1. Кликните на **OK**. Резултати ће бити приказани у новом радном листу под називом Subtract\_Sheet1.

# Приказ Венових дијаграма на основу интервалних геномских података

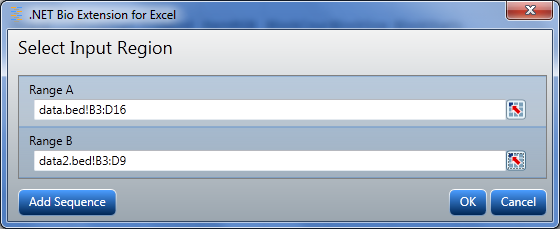
Опцијом Venn Diagram (Венов дијаграм)**,** можете направити дво- или тро-обласне Венове дијаграме на основу интервалних геномских података у формату BED, што омогућује визуелизацију односа међу областима, као и визуелизацију преклапања података.

Напомена: Venn Diagram захтијеве NodeXL шаблон за Excel 2007 или Excel 2010, доступан на <http://www.codeplex.com/NodeXL>.

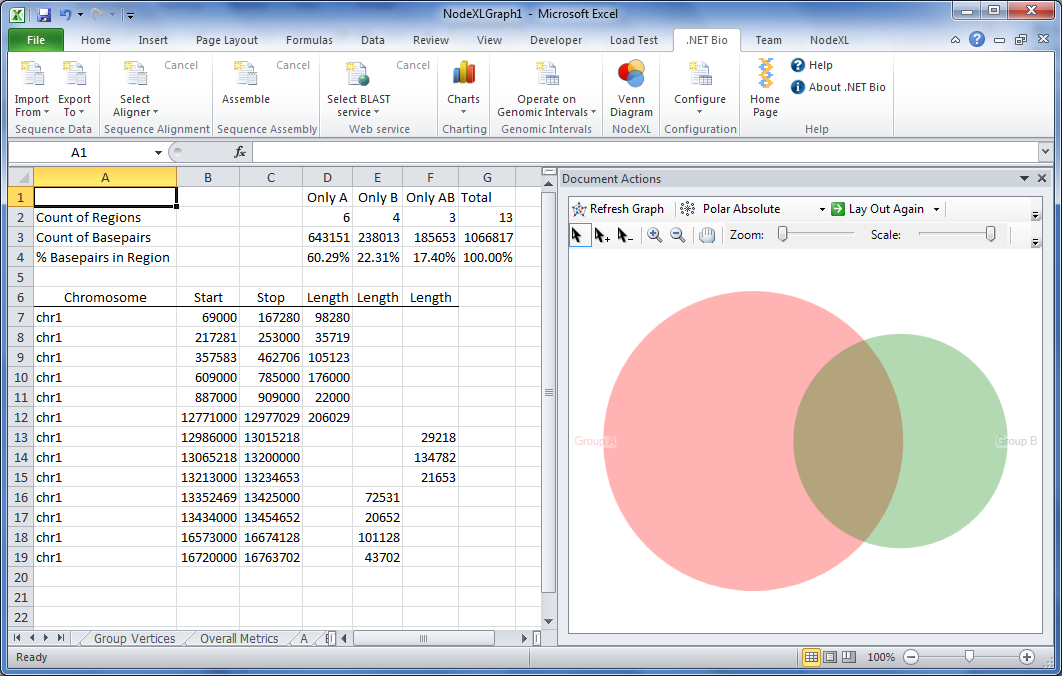
За формирање дво-обласног Веновог дијаграма

1. Кликните на Import From и одаберите BED.
2. Означите једну или више датотека и кликните на Open.
3. Означите радни лист и кликните на Venn Diagram.
4. Користите прозор Select Input Sequence Ranges, као што је приказано на сљедећој слици, да изаберете два опсега базних парова.

ВАЖНО: Опсези базних парова се морају преклапати. Нпр. коришћење Chr1 за геноме горила и људи.

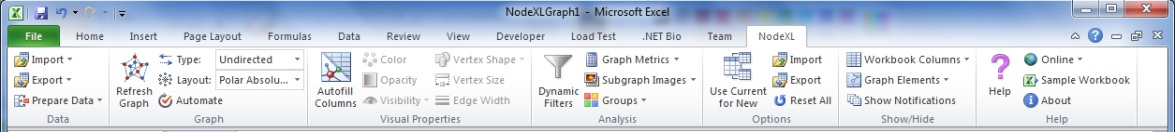
  
Дијалошко окно Select Input Sequence Ranges

1. Резултујући Веновог дијаграм је приказан у новој радној свесци под називом NodeXLGraph1, као што се може видјети на сљедећој слици.

  
Венов дијаграм генома Chr1код људи и горила

1. Кликните на изборник NodeXL да бисте видјели ленту NodeXL, као што је приказано на сљедећој слици.

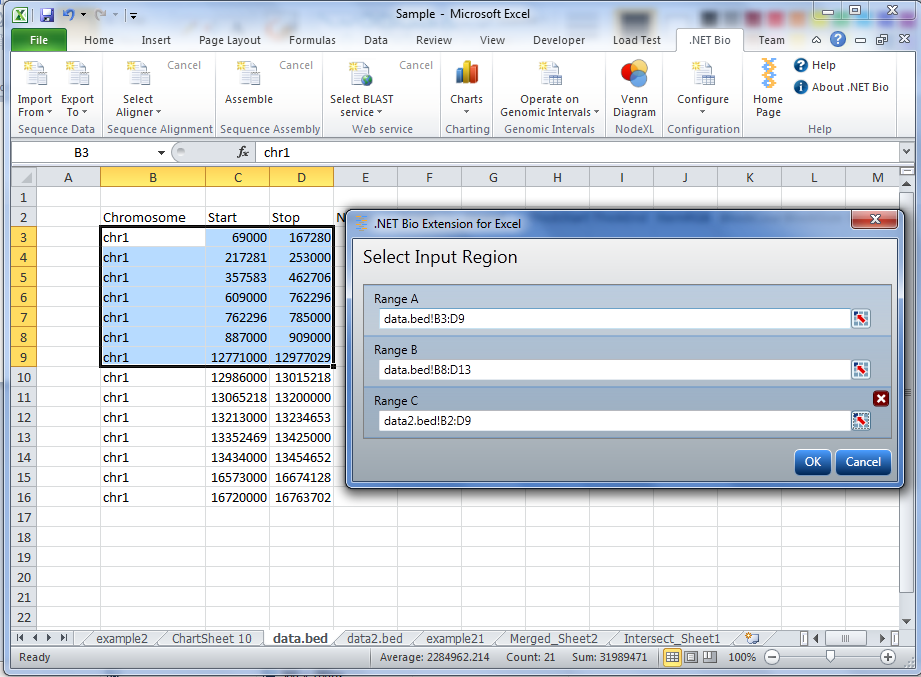
Документација ових наредби се налази на адреси http://www.codeplex.com/NodeXL

  
NodeXL лента

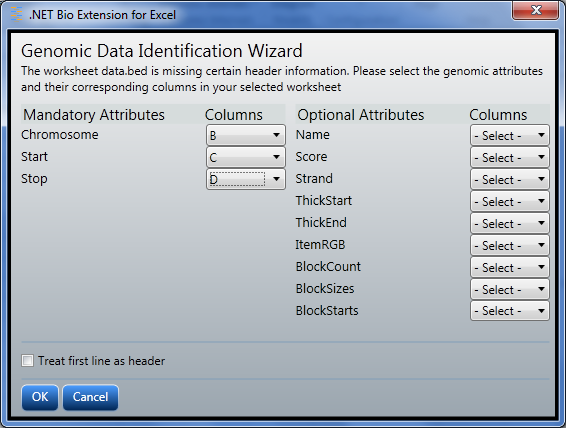
За формирање тро-обласног Веновог дијаграма

1. Кликните на Import From и одаберите BED.
2. Означите једну или више датотека и кликните на Open.
3. Означите радни лист и кликните на Venn Diagram.
4. Користећи прозор Select Input Sequence Ranges, као што је приказано на сљедећој слици, изаберите три опсега базних парова и кликните на **OK**.

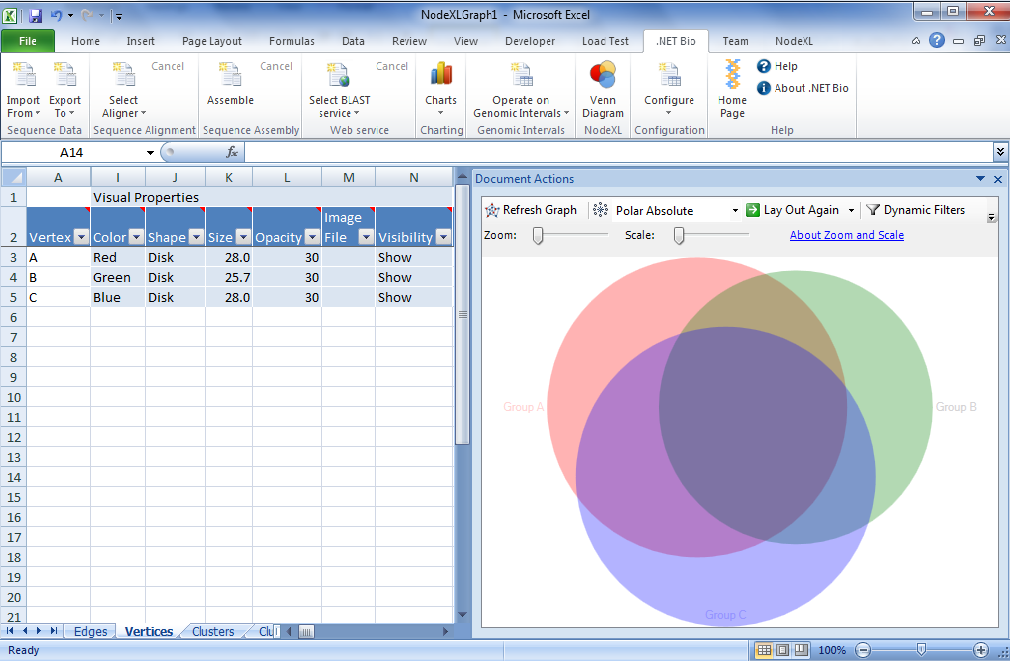
У овом примјеру, изабрана три опсега базних парова се преклапају.

 Три опсега базних парова који се преклапају

1. Подесите вриједност колона помоћу чаробњака Genomic Data Identification Wizard, као што је приказано на сљедећој слици, и кликните на **OK**.

  
Чаробњак за идентификацију геномских података

1. Резултати Веновог дијаграма су приказани у новој радној свесци под називом NodeXLGraph1, као што можемо видјети на сљедећој слици.

Тро-обласни Венов дијаграм

# Промјена конфигурационих опција

.NET Bio Extension има двије конфигурационе опције:

* број колона-оквира за податке о геномским секвенцама;

Мијења се начин на који Excel показује податке о секвенцама. Подразумјевани оквир је 80, а максимална вриједност је 16 000.

* шема за бојење молекула секвенце;

Сваком молекулу се може придружити боја ради јасније представе секвенце. Само пет молекула, подразумјевано, има одређену боју: A, T, C, G, и U.

За конфигурисање колоне-оквира

1. Кликните на Configure на ленти .NET Bio.
2. Кликните на колону Sequence Data Wrapround (Оквир података секвенце).
3. Унесите нову вриједност у поље Enter the maximum number of columns (Унесите максималан број колона).

За конфигурисање шеме боја

1. Кликните на Configure на ленти .NET Bio.
2. Кликните на Change Color Scheme for Molecules (Измјена шеме за бојење молекула)**.**
3. Кликните на дугме Change Color (Промијени боју) у дијалошком окну Configure Color**.**
4. Означите боју у дијалошком окну Format Cells (Формати ћелија)и кликните на **OK**.
5. Кликните на **OK** у прозору Configure Color да снимите промјене.

# Додатак A: Подржане секвенце и формати датотека

У овом додатку су описани подржани формати .NET Bio радног оквира, са одговарајућим хипервезама ка референцама за више информација.

### FASTA: Секвенце података

Једноставан текстовни формат за представљање пептидних или нуклеотидних ланаца, а да је при том анализа и манипулација секвенцама доста лака помоћу скрипти написаних на језику као што је Iron Python.

Формат је у ствари низ линија од, најчешће, 80 слова по линији, али никако не више од 120.

Технички опис

http://www.ncbi.nlm.nih.gov/blast/fasta.shtml

Извори

Википедијина страница <http://en.wikipedia.org/wiki/FASTA_format> – садржи општи преглед, хипервезе ка конверторима формата и још неке битне референце.

### FASTQ: Квалитативне секвенце података

Текстовни формат који похрањује биолошке секвенце и Phred quality бодове у једну датотеку. Често се сматра *de facto* стандардом за похрану хеуристичких и бодовних података високо-пропусног анализатора (High-Troughput Computing).

Формат је одређен са четири линије датотеке које чине слог (слог одговара једној геномској нисци).

Уобичајне екстензије датотека су: .fq, .fastq, .txt.

Технички опис

FASTQ формат спецификација  
<http://maq.sourceforge.net/fastq.shtml>

Извори

Википедијина страница <http://en.wikipedia.org/wiki/FASTQ_format> – садржи општи преглед, хипервезе ка конверторима формата и још неке битне референце.

### GenBank: Формат за базе података нуклеотидних секвенци

Flat-file формат који описује нуклеотиде и секвенце нуклеотида из GenBank-а – базе података са слободним приступом.

Технички опис

“Chapter 1, GenBank: The Nucleotide Sequence Database,” Ilene Mizrachi; *NCBI Handbook*, 2007

<http://www.ncbi.nlm.nih.gov/books/bookres.fcgi/handbook/ch1.pdf>

Извори

Веб-сајт NCBI базе података за опште информације о GenBank-у <http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide>

Википедијина страница <http://en.wikipedia.org/wiki/GenBank> – садржи општи преглед, хипервезе ка конверторима формата и још неке битне референце.

### GFF: Generic Feature Format

Линијски формат који користи tab-размаке за раздвајање података. Намијењен је за представљања слогова у геномској бази података. GFF слог представља субсеквенцу неке биолошке секвенце, као што је то ген или протеинска секвенца, истовремено дозвољавајући „умјерено детаљна“ образложења.

Екстензија за овај тип датотека је .gff.

Раније спецификације су преводиле акроним као Gene-Finding Format.

Технички опис

Садашња верзија је n2. Формат су првобитно осмислили Richard Durbin и David Haussler, а посљедња верзија садржи измјене које су предложили Lincoln Stein, Suzanna Lewis, Anders Krogh и други.

<http://www.sanger.ac.uk/resources/software/gff/spec.html>

Извори

Сајт института Wellcome Trust Sanger за општи преглед   
<http://www.sanger.ac.uk/Software/formats/GFF/GFF_Spec.shtml>

UCSC сајт пројекта Encode, такође за општи преглед  
<http://genome.ucsc.edu/goldenPath/help/customTrack.html#GFF>

### Browser Extensible Data (BED) Format

Формат BED обезбјеђује флексибилан начин за дефинисање линија које приказују у дијелу за образложење. BED се састоји од главне датотеке са пољима одвојеним tab-размацима и слоговима раздвојеним простим празнинама.

Екстензија за овај тип датотека је .bed.

FAQ

Browser Extensible Data (BED) Format FAQ  
http://genome.ucsc.edu/FAQ/FAQformat.html#format1

Извори

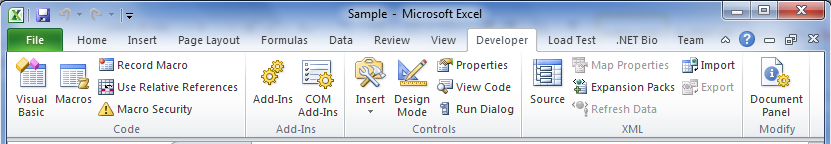
UCSS-ов сајт са општим прегледом базе података Genome Browser Database   
http://users.soe.ucsc.edu/~kent/gbd.html#BED

# Додатак Б: Одобравање макроа

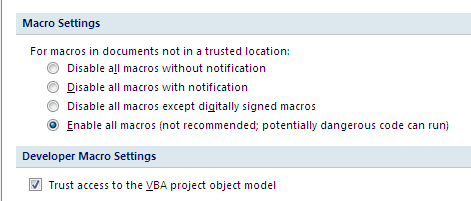
Да бисте користили Charting функцију, морате одобрити макрое у Excel-у и додати Excel-макро радним свескама које користите у Biology Extension-у. Назив макроа је DisplayDNASequenceDistribution.bas, и долази инсталиран уз .NET Bio радни оквир.

Одобравање макроа у Excel-у

1. Уз текућу отворену радну свеску, кликните на Office Button у Excel-у и кликните на Excel Options.
2. Кликните на Show Developer tab in the Ribbon, а онда на OK.
3. Кликните на картицу Developer на ленти, као што је приказано на сљедећој слици.

  
Developer картица

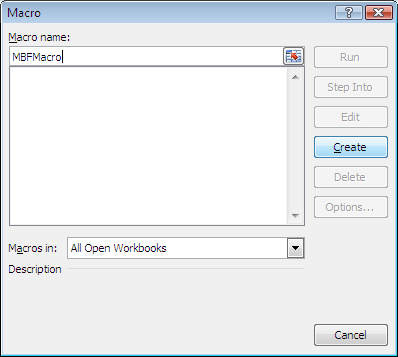
1. Кликните на Macro Security у ленти Developer.
2. Кликните на Macro Settings у прозору Trust Center.
3. Кликните на Enable all macros и означитеTrust access to the VBA project object model као што је приказано на сљедећој слици.

  
Macro Settings

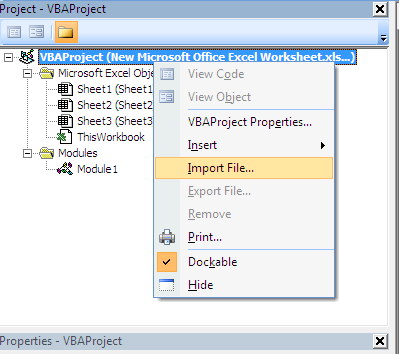
1. Кликните на **OK**.
2. Затворите и опет отворите радну свеску да би промјене ступиле на снагу.

За додавање графикон-макроа

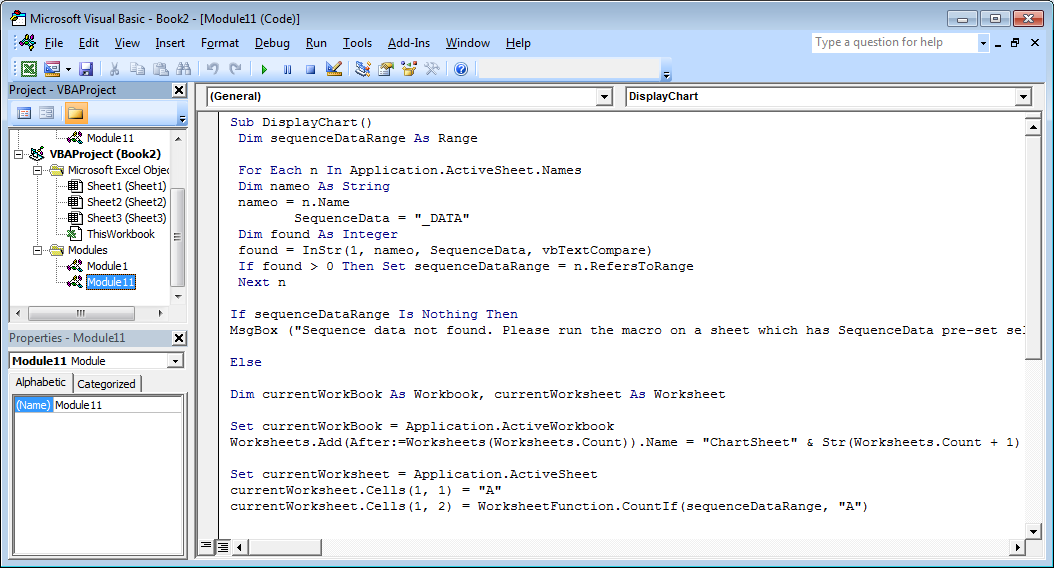
1. Кликните на картицу Developer на траци и кликните на Macros.
2. Унесите има макроа као што је „BioMacro” у поље Macro name у прозору Macro, као што је приказано на слици и кликните на Create.

  
Макро прозор

1. Десни клик на VBAProject у прозору Microsoft Visual Basic, а онда кликните на Import File… као што је приказано на слици.

  
Наредба Import File…

1. Идите у C:\Program Files (x86)\.NET Bio\1.0\Tools\.NET Bio Extension for Excel. Означите DisplayDNASequenceDistribution.bas и кликните на Open**.**
2. Двоструким кликом миша на Module 11 да се прикаже макро, као што је приказано на слици.

  
Mакро DisplayChart

1. Кликните Save (Сачувај) и сачувајте документ типа Excel Macro-Enabled Workbook (\*.xlsm).

Сада, када сте допустили макрое у Excel-у и додали марко DisplayDNASequence-Distribution.bas, можете користити Charting функцију у Biology Extension-у.