**Comment [d2]:** In total we have 14 n-domains architecture in whole dataset. There sequence count is as given below:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| n-Domains | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 9 | 10 | 11 | 12 | 13 | 15 | 21 |
| Sequence Count | 160 | 22 | 51 | 11 | 76 | 2 | 15 | 6 | 2 | 17 | 1 | 4 | 3 | 1 |

**Comment [d3]:** I can find out the new sequences and add the corresponding attributes. However this would change all the stats and every analysis has to be redone. And this will happen every time when new sequences are reported. Therefore I feel we should do reporting with the base data and in future new records can be added through Information System.

**Comment [d4]:** Total IRDs = 698. IRD’s found at least in one sequence are 694. Therefore total number of IRD’s NOT found in any sequence is: 4 and they are: IRD-71; IRD-148; IRD-331; IRD-688

**Comment [d6] –[d8] :** Refer Excel: dom\_rcl\_linker.csv

|  |  |  |
| --- | --- | --- |
| Type-1 | H-L Type | 238 |
| Type-2 | L-H Type | 256 |
| Type-3 | H+L type | 201 |

**Comment [d11]:** Refer Excel - d11\_org\_seq\_freq.csv

**Comment [d12]:** Refer Excel - d12\_d13\_organism\_n-domain\_distribution.csv

* Solanaceae plants consisted of following n-domains :

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| n-Domains | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 9 | 10 | 11 | 12 | 13 | 15 | 21 |

* Single Domain IRD’s in Solanaceae family= 61 sequences, distributed across 10 species

Single Domain IRD’s in non Solanaceae family= 99 sequences , distributed across 60 species

**Comment [d14]:** Majority of IRD’s belong to: **Capsicum annuum (Bell pepper).** Refer Excel: *d14\_organism\_domain\_distribution.csv* and *d14\_organism\_domain\_uniqueDistribution.csv*

**Comment [d15- d19]:** Refer Excel: d15\_d19\_UniqueDomainsFrequency\_sequenceWise.xlsx

**Comment [d20]:**

|  |  |  |
| --- | --- | --- |
| Total Linkers present | Type1 | Type2 |
| 75 | 24 | 51 |

* Refer Excel: d20\_LinkerSequenceDiversityAndVariation.xlsx for Linkers Sequence Diversity and variation.
* Refer Excel: d20\_LinkerOrganismDistribution.csv

|  |  |  |
| --- | --- | --- |
| **Linker Type** | **Family** | **Occurrences** |
| Type 1 | Capsicum annuum (Bell pepper) | 226 |
|  | Nicotiana tabacum (Common tobacco) | 49 |
| Type 2 | Capsicum annuum (Bell pepper) | 164 |
|  | Nicotiana tabacum (Common tobacco) | 61 |

**Comment [d21-d22]:**

Refer Excel: d21\_d22\_LinkerDistributionAcrossOrganism.csv

**RCL distribution and specificity**

**Comment [d23]:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Total IRD’s present** | **Total RCL present** | **CPRNC** | **CTLNC** | **CPRYC** |
| 695 | 63 | 578 | 113 | 94 |

Also for RCL Occurrences in species Refer Excel: d24\_RclOccurrenceInSpecies.csv

**Comment [d24]:** For Solanaceae, Non Solanaceae diversity Refer Excel: d24\_SolanaceaeDiversityOfRcl.cs

**Comment [d25]:** For Organism Diversity of RCL Refer: d25\_OrganismDiversityOfRCL.csv

**Comment [d26-d28]:**

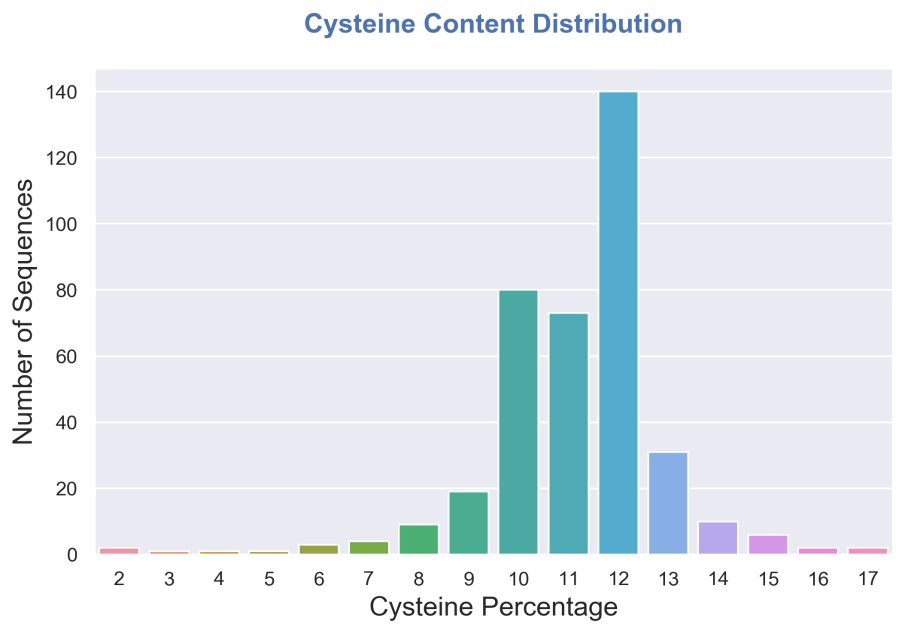
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Total RCL** | **Trypsin** | **Chymotrypsin** | **Elastase** | **Unknown** |
| 63 | 15 | 20 | 3 | 25 |

**Comment [d29]:**

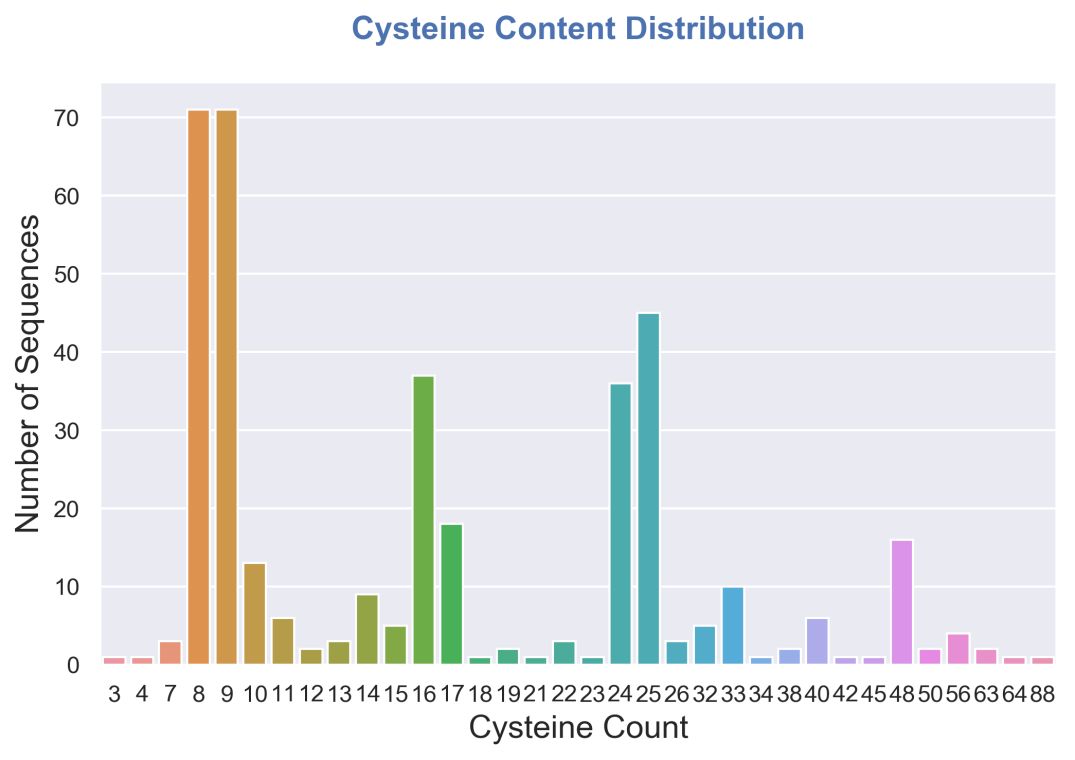
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Total IRD’s** | **Trypsin** | **Chymotrypsin** | **Elastase** | **Unknown** |
| 695 | 396 | 177 | 12 | 110 |

**Comment [d30]:**  
For distribution of Target Specificity across species Refer Excel: d30\_targetSpecificityAcrossSpecies.csv

Comment [d31- d32]: **Disulphide bond variations in the Pin-II family**Refer Excel: CysteineCountInDomains.csv, CysteineCountInSequences.csv  
Refer Graph: d31\_CysteineCountDistributionInSequences.png, d32\_CysteineCountDistributionInDomains.png

Fig 3a was for Pin-II Sequences and NOT for Domain Sequences. And X-axis was Cysteine Percentage  


Also the interpretation for Fig 3a that “*Analysis of number of Cysteines across the Pin-II family indicated that most of the Pin-II PI sequences contain 12 Cys residues, followed by sequences with 10 Cys residues* “ is little bit incorrect since the X-axis of the Fig 3a graph depicts **Cysteine Percentage** and NOT **Cysteine Count.**Following figure will give you the actual picture if you want to consider **Cysteine Count**.



Also if you want to see the **Cysteine Count** distribution corresponding to **Domain** sequences then following is the graph.  
