PLAZI

TAKING CARE OF FREEDOM

https://plazi.org/

Mobilise Action
Mobilising Data, Policies and Experts in Scientific Collections

TreatmentBank

Biodiversity Literature Repository

eBioDiv

CERN

SIBILS
SWISS INSTITUTE OF BIOINFORMATICS LITERATURE SERVICES

swissuniversities

101007492

BiCIKL

ARCADIA

Plazi: QC Tutorial

Julia Giora
Jonas Blanco
Donat Agosti

Quality Control

➤ What is the Plazi Quality Control?

- Liberation of treatment information:
    - Template
    - Individual extraction

- GGI
    - Mark sections, figures, taxon names, parse material citation atributes…
    - Boundaries
    - Paper structure

➤ What is the Plazi Quality Control?

- GGI
    - Mark sections, figures, taxon names, parse material citation atributes…
    - Boundaries
    - Paper structure

➤ What is the Plazi Quality Control?

- Machine detectable: prevent wrong information to reach the repositories
    - **Blockers**
    - **Criticals**
    - Majors
    - Minors

➤ What is the Plazi Quality Control?

- To detect and correct the extraction errors!

➤ What is the Plazi Quality Control?

- Granularity
    - Quality level
    - High
    - Low
    - Automate

➤ What is the Plazi Quality Control?

- **Granularity**
  - **Quality level**

High-Level

> Applied to: **European Journal of Taxonomy**

QC Protocol: Usually all. For EJT 2020's, Blocker+ (meaning ALL Blockers and Criticals related to treatment structure).

- Treatment boundaries;
- Images, images citation, captions;
- Tables, tables citation, captions;
- All materialsCitations with all annotations included;
- All treatmentCitations with correct attributes;
- All bibRefs correctly annotated;

➤ What is the Plazi Quality Control?

- **Granularity**
  - **Quality level**

Low-Level

> Applied to: **MNHN Journals, such as Zoosystema, Adansonia, Anthropozoologica and Geodiversitas. Also heavily used to most processed articles from Feb/2020 to Apr/2020.

QC Protocol: Blocker+.

- Treatment boundaries;
- Images, images citation, captions;
- Tables, tables citation, captions;
- All materialsCitations boundaries, but annotating details only for Holotypes;
- All treatmentCitations with correct attributes;
- All bibRefs correctly annotated;

➤ What is the Plazi Quality Control?

Automate-Level

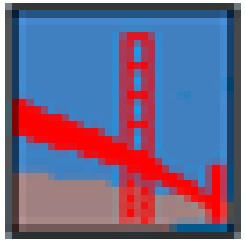
> Adopted in May/2020 upon Donat's request to process **EVERYTHING** except EJT 2020's (high-level), EJT backlog () and MNHN journals at this moment.

QC Protocol: None.

- Open IMFs that did not generate DwC file, fix the problem, count stats;
- Open 10% of the IMFs,
- Stats being count:
  - missing treatments
  - spurious treatments
  - missing keys
  - missing images
  - number of duplicated images
  - missing tables
  - missing bibRefs
  - proportion of missing materialsCitation
  - proportion of missing treatmentsCitation

➤ How to proceed a Quality Control?

- Files from GG Server
- Open GGI
- File
  - Load Document From GG Server
  - Ok - Username and password
  - Insert paper UUID



➤ How to proceed a Quality Control?

- Set a stopwatch to register the time spent on each QC
- Set the quality level
  - Usually: Automate level

  - Manually check page structure
    - Word flow
    - Text stream type
    - Paragraphs
    - Blocks
    - Column
    - Images

  - Manually check document structure
    - Bibliographic reference
    - Caption
    - Taxonomic names
    - Material Citation
    - Treatment
    - Subsubsection

TAXON 60 (4) • August 2011: 1138–1153

A new classification of the South African endemic family Bruniaceae based on molecular and morphological data

Regine Classen-Bockhoff,[1] Edward G.H. Oliver,[2] Anthony V. Hall[3] & Marcus Quint[1]

1 *Institut für Spezielle Botanik, Johannes Gutenberg-Universität Mainz, Germany*
2 *Department of Botany & Zoology, Stellenbosch University, South Africa*
3 *Bolus Herbarium, University of Cape Town, South Africa*
Authors for correspondence: *Regine Classen-Bockhoff, classenb@uni-mainz.de; Edward G.H. Oliver, eoliver@sunlac.za*

Abstract The classification of Bruniaceae is reviewed based on molecular (*matK*, ITS, *rbcL*) and morphological data, and the tribes and genera are redefined as monophyletic, morphologically diagnosable lineages. The family is subdivided into three tribes and six genera with 81 accepted species: Linconieae (*Linconia*), Audouinieae (*Audounia* [incl. *Titmannia*], *Thamnea* [incl. *Pseudobaeckea teres*]), and Brunieae (*Berzelia* [incl. *Brunia* p.p. *Staavia* incl. *Raspalia staavioides*], and *Brunia* [incl. *Nebelia*, *Pseudobaeckea* p.p., *Raspalia* p.p., *Lonchostoma*, *Mniothamnea*]). A key for the new classification, a short description of each genus and an updated nomenclature of all the species are provided. Two new tribes are described and thirty-eight new combinations and names are provided.

Keywords *Audouinia*; *Berzelia*; *Brunia*; Bruniaceae; *Linconia*; *Staavia*; taxonomy; *Thamnea*

Supplementary Material The alignment is available in the Supplementary Data section of the online version of this article (http://www.ingentaconnect.com/content/iapt/tax).

INTRODUCTION

Bruniaceae R. Br. ex DC. are a small family of flowering plants endemic to South Africa. It represents one of the 33 "Cape floral clades" (Linder, 2003) currently under investigation for the reconstruction of the history of the Cape Floristic

(Leinfellner, 1964a, b; Quint & Classen-Bockhoff, 2006b), phytochemistry (Scott, 1999), growth form and inflorescence morphology (Classen-Bockhoff, 2000), phylogeny and dating of speciation events (Quint, 2004; Quint & Classen-Bockhoff, 2006a, 2008).

The taxonomy at species level was slightly revised by

Display Control
word
Regions, Blocks, etc.
Show All | Hide All
block
column
image
line
paragraph
region
table
tableCell
tableCol
tableRow
Annotations
Show All | Hide All
author
bibCitation
bibRef
bibRefCitation
bookContentInfo
caption
collectingCountry
collectingDate
collectingMunicipality
collectingRegion
collectionCode

Authors for correspondence: Regine Classen-Bockhoff; classenb@uni-mainz.de; Edward G.H. Oliver; eoliver@sunlac.za

Abstract The classification of Bruniaceae is reviewed based on molecular (*matK*, ITS, *rbcL*) and morphological data, and the tribes and genera are redefined as monophyletic morphologically diagnosable lineages. The family is subdivided into three tribes and six genera with 81 accepted species: Linconieae (*Linconia*), Audouinieae (*Audounia* incl. *Tittmannia*), *Thamnea* (incl. *Pseudobaeckea teres*), and Brunieae (*Berzelia* incl. *Brunia* p.p., *Staavia* incl. *Raspalia staavioides*, and *Brunia* incl. *Nebelia*, *Pseudobaeckea* p.p., *Raspalia* p.p., *Lonchostoma*, *Mniothamnea*). A key for the new classification, a short description of each genus and an updated nomenclature of all the species are provided. Two new tribes are described and thirty-eight new combinations and names are provided.

Keywords *Audouinia*; *Berzelia*; *Brunia*; Bruniaceae; *Linconia*; *Staavia*; taxonomy; *Thamnea*

Supplementary Material The alignment is available in the Supplementary Data section of the online version of this article (http://www.ingentaconnect.com/content/iapt/tax)

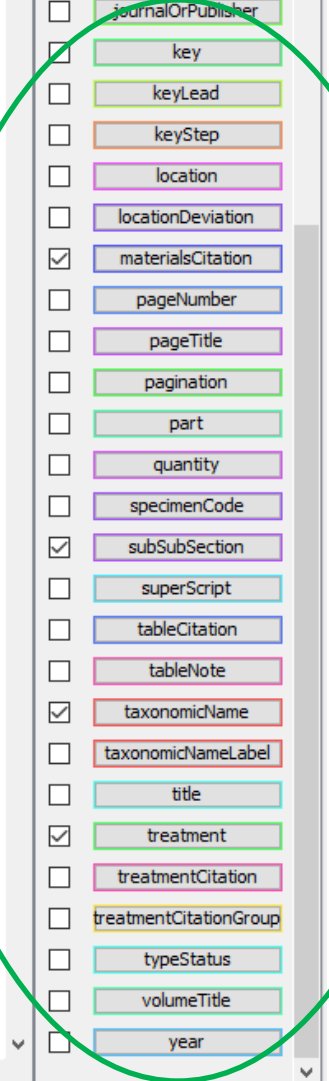
INTRODUCTION

Bruniaceae R. Br. ex DC. are a small family of flowering plants endemic to South Africa. It represents one of the 33 "Cape floral clades" (Linder 2003) currently under investigation for the reconstruction of the history of the Cape Floristic Region. It also has a key position in the understanding of the evolutionary history of the campanulid clade (Winkworth & al. 2008). Given the importance of the family we aim to provide a revised generic classification of the family based on recent phylogenetic analysis (Quint & Classen-Bockhoff, 2006a).

The taxonomic history of Bruniaceae dates back to Brown (1818) who established the family. The last revision of the family is that of Pillans (1947) who based his work on that of Brongniart (1826), Sonder (1862), Dümmer (1912), and Niedenzu & Harms (1930). He distinguished 12 genera and 75 species using number of styles, locules and ovules, fusion of floral parts, inflorescence structure, anther shape, and stamen length as the main diagnostic characters. His classification has underlain all subsequent morphological and systematic work of the family (e.g., Takhtajan 1987; Carlquist 1978, 1991; Hall 1988; Classen-Bockhoff 2000). Only Dyer (1975), citing an unpublished classification provided by E. Powrie, included *Nebelia* in *Brunia*, thus reducing the number of genera to eleven (Leinfellner 1964a, b; Quint & Classen-Bockhoff, 2006b), phytochemistry (Scott, 1999), growth form and inflorescence morphology (Classen-Bockhoff, 2000), phylogeny and dating of speciation events (Quint, 2004; Quint & Classen-Bockhoff, 2006a, 2008).

The taxonomy at species level was slightly revised by Powrie (1969a) following research on type material, and by Goldblatt & Manning (2000). Five new species have been described since Pillans' revision: *Lonchostoma esterhuyseniae* (Strid, 1968), *Tittmannia esterhuyseniae* (Powrie, 1969b), *Linconia ericoides* (Oliver & Oliver, 1999), *Brunia compacta* and *Thamnea matroosbergensis* (Hall & al., 2010). The obvious misinterpretation of *Pseudobaeckea teres* (Hall, 1988; Carlquist, 1991; Classen-Bockhoff, 2000) and increased knowledge of inter- and intrapopulation variation, mainly due to the substantial herbarium collections of E. Esterhuysen, prompted Hall (2000, 2001) to undertake a new revision of the family and to propose a new generic classification. Many of his results corresponded with an independent molecular analysis of the family (Quint & Classen-Bockhoff, 2006a). The primary incongruence between the two studies is the identification of *Raspalia* as a polyphyletic taxon in the molecular analysis (Fig. 1).

We present here a revised generic classification based primarily on the molecular and morphological data presented by



➤ How to proceed a Quality Control?

- Check paper extructure
  - Check treatments boundaries
  - Paragraphs organization
  - Misidentified material citation

BRUNIACEAE R. Br. ex DC.

Key to the tribes genera and subgenera (see Figs 2–4)

1. Anthers apically ending in a conspicuous sterile tip (Linconieae) ..... 1. ***Linconia***
1. Anther apex without a sterile tip 2
2. Thecae connate along their entire length, anthers not versatile; stomata often surrounded by cuticular rim (Audouinieae) 3
2. Thecae free at their base, anthers versatile, stomata never surrounded by cuticular rim (Brunieae) 4
3. Flowers arranged in spike-like clusters, each flower on a bracteate short-shoot; flower pedicels present, stomata surrounded by cuticular rim ..... 2. ***Audouinia***
3. Flowers solitary, each on a leafy shoot or bracteate short shoot (only in *Thamnea teres*); flower pedicels absent (except *T. subtilis*), stomata rarely surrounded by cuticular rim (only in *T. teres*) ..... 3. ***Thamnea***
4. Styles 2, united; inflorescences with basipetal flowering sequence ..... 4. ***Staavia***
4. Styles 1 or 2, free; flowering sequence not as above 5
5. Pollen 3-colporate; stamens exserted; inflorescences laterally arranged ..... 5. ***Berzelia***
5. Pollen 5- to 10-colporate; stamens shorter than corolla (except *Brunia* subg. *Brunia*); terminal inflorescence usually present ..... 6. ***Brunia*** → 6 for subgenera
6. Stamens exserted; petals with 2 narrowly associated ridges ..... 6.1 ***B.*** subg. ***Brunia***
6. Stamens included; petals different from above 7
7. Petal bulges forming a thick cushion with vertical subdivision; petals free, less than 5 mm long 8
7. Petal bulges reduced or absent; petals postgenitally fused with filaments, more than 5 mm long ..... 6.5 ***B.*** subg. ***Lonchostoma***
8. Leaves erect-spreading ..... 6.2 ***B.*** subg. ***Pseudobaeckea***
8. Leaves appressed to partly imbricate 9
9. Inflorescences usually larger than 5 mm in diameter, leaves hairy (except *B. angulata*, *B. microphylla*) ..... 6.3 ***B.*** subg. ***Raspalia***
9. Inflorescences usually 2–4 mm in diameter or flowers solitary (*B. bullata*, *B. callunoides*); leaves glabrous ..... 6.4 ***B.*** subg. ***Mniothamnea***

For a full synonymy and descriptions of the species consult Pillans (1947). To help locate the species a comparison of Pillans's classification and the new one is given in the Appendix. A paper with more complete nomenclature and typification of the species and notes thereon will be published separately

The 16 species not included in the molecular analysis are added to the respective genera based on their corresponding morphological data. The species are *Audouinia hispida*, *Berzelia commutata*, *B. dregeana*, *B. squarrosa*, *Brunia barnardii*, *B. compacta*, *B. laevis*, *B. palustris*, *B. pillansii*, *B. schlechteri*, *B. tulbaghensis*, *Staavia trichotoma*, *S. staavioides*, *Thamnea depressa*, *T. gracilis*, and *T. matroosbergensis*.

[Tr. **Linconieae** Quint & Class.-Bockh. **tr. nov.**, floribus in fasciculis brachyblastorum bracteatorum antheris apice sterili conspicuo, stomatis crista cuticulari circumcinctis. – Type *Linconia* L.]

Anthers ending apically in a conspicuous sterile tip; stomata surrounded by a cuticular rim.

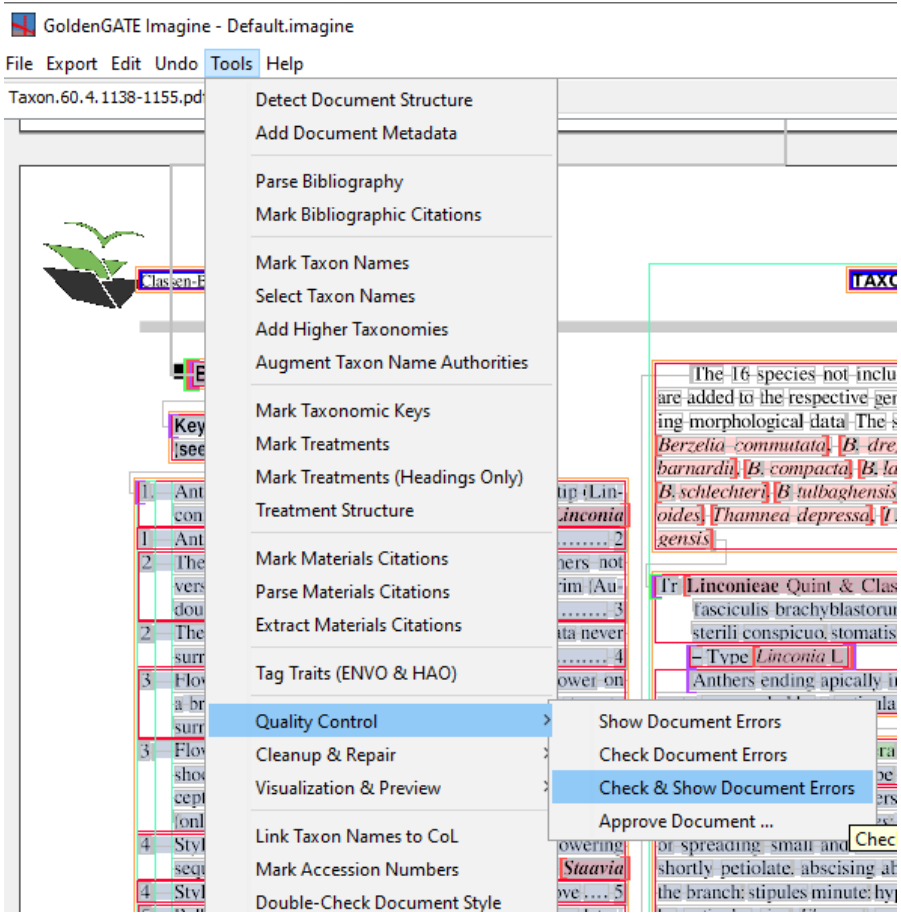
1. ***Linconia*** L., Mant. Pl. Altera 216. 1771; Pillans in J. S. African Bot. 13: 130. 1947. – Type *L. alopecuroidea* L.

Growth form: erect undershrubs or rock crevice shrublets, lignotubers present. *Leaves*: imbricate (*L. alopecuroidea*) or spreading, small and ericoid (*L. cuspidata*, *L. ericoides*), shortly petiolate, abscising above petiole which remains on the branch; stipules minute; hypostomatic, stomata surrounded by cuticular rim. *Flowers*: pedicellate, terminal on bracteate short-shoots, arranged in spike- or head-like clusters at top of vegetative branches that continue to grow after flowering, usually with 4 bracts per flower. *Petals*: bright pink (*L. alopecuroidea*, *L. ericoides*) or creamy white (*L. cuspidata*), with two ventral thick ridges forming a "V", the apex approaching each other towards base of petal, fused basally in a low median ridge. *Scent*: lacking. *Stamens*: included, anthers sagittate (distal ends of thecae clearly diverging and apical ends fused) with pale, sterile tip, blue (white in *L. cuspidata*). *Pollen*: 3-colporate. *Ovary*: half to two-thirds inferior, 2-locular (rarely 3-locular in *L. alopecuroidea*) with 2 ovules per locule. *Styles*: two, free. *Fruit*: 1- or 2-seeded dry dehiscent. *Habitat*: relatively dry rock crevices (*L. cuspidata*, *L. ericoides*), or in moist fynbos (fine-leaved shrub vegetation) in partial shade of vegetation on southern slopes (*L. alopecuroidea*), at medium altitudes of 500–600 m (*L. alopecuroidea*, *L. ericoides*), or at high altitudes of 900–1500 m (*L. cuspidata*). *Flowering time*: September to November. *Species*: 3 spp.; south-western region of Western Cape extending to Heidelberg (Humansdorp vide Pillans, 1947).

Linconia alopecuroidea L., Mant. Pl. Altera: 216. 1771; Pillans in J. S. African Bot. 13: 131. 1947. – Lectotype: Herb. Linn. 323.1 (LINN, lecto. designated by Hall in Regnum Veg. 127: 61. 1993).

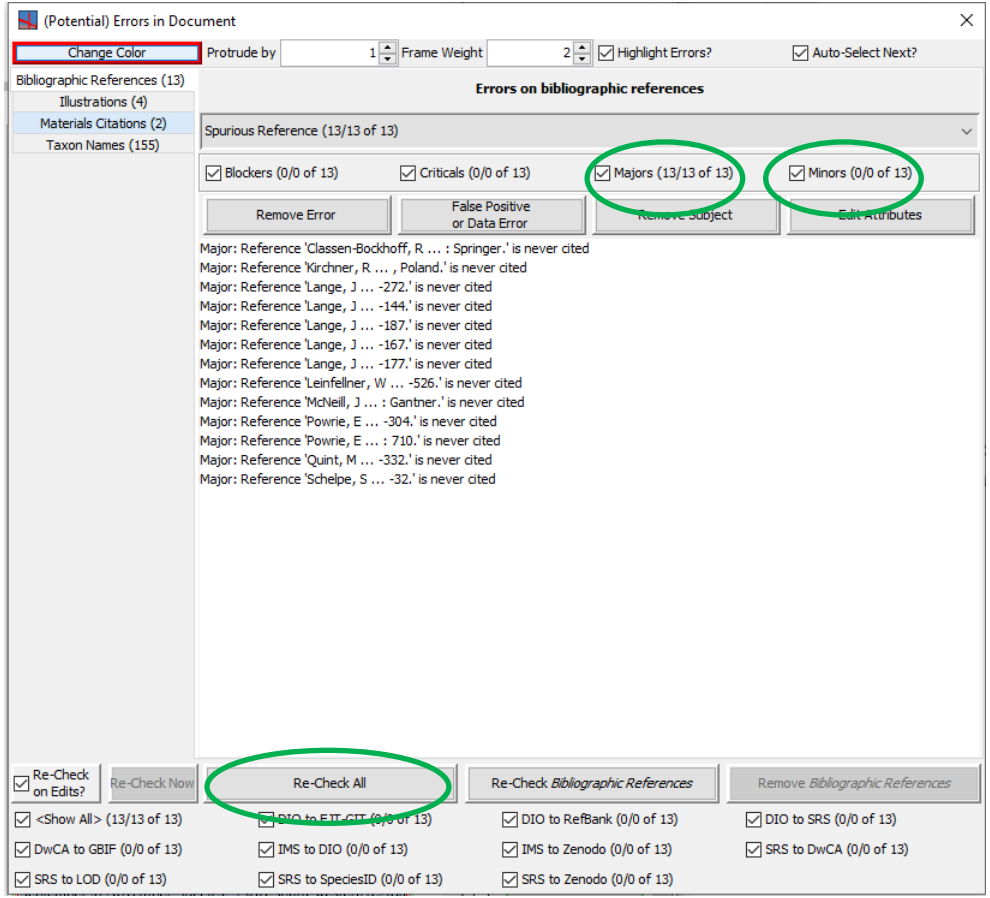
➤ How to proceed a Quality Control?

- Tools
- Quality Control
- Check and Show Document Errors



➤ How to proceed a Quality Control?

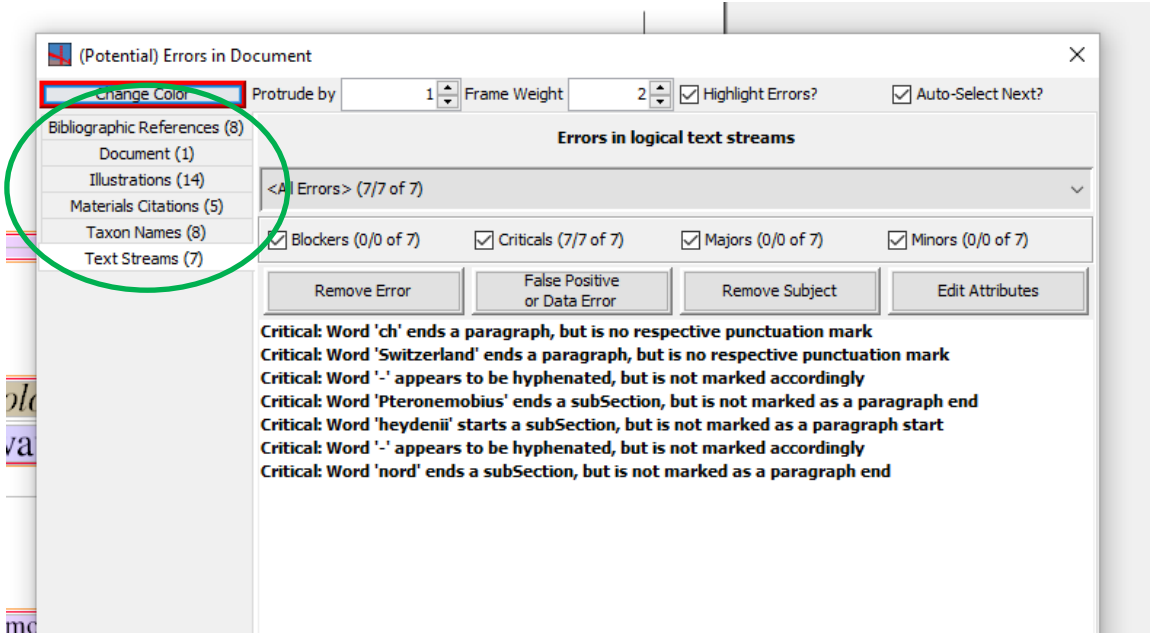
- Check and Show Document Errors
  - Unmark Major Errors and Minor Errors
  - Re-check All



➤ How to proceed a Quality Control?

- Order for an easier errors analysis
  - Text
  - Illustrations
  - Treatments
  - Material Citations
  - Bibliographic References

- Save Documento to GoldenGate Server



Checklist to Avoid and/or Correct Errors

Some processing mistakes, or lack of information will always result in errors that impede data transference to the repositories.

Document Metadata:

- Identifiers (all CRITICAL errors):
  - check if the document has a DOI
  - check if the document has a Zenodo deposition ID

- Bibliographic Metadata (all BLOCKER errors unless indicated otherwise):
  - check if the document has at least one author
  - check if the document has a title
  - check if the document has a year
  - check if the document has a publication date (CRITICAL error if absent)
  - check if the document has a pagination
  - check if the document has a journal name
  - check if the document has a volume number

Text Streams:

- Paragraph Starts:
  - check if any paragraphs start with a lower case word longer than one character (CRITICAL error on paragraph if they do)

- Paragraph Ends:
  - check if each main text paragraph (outside a key or heading) ends with a period, colon, or exclamation or question mark (CRITICAL error on paragraph if they don't)

Illustrations:

- Captions:
  - check if captions have attributes linking it to image, graphics, or table (CRITICAL error if not)
  - check if captions assigned to images or graphics have an HTTP URI (MAJOR error if not)
  - check if captions are referenced by at least one figureCitation or tableCitation (MAJOR error on caption if not)
  - check if paragraphs starting with an emphasized caption start keyword ('Figure', 'Table', etc.) are marked as captions (CRITICAL error on paragraph if not)

- Caption Citations:
  - check if figureCitations and tableCitations are linked to captions (CRITICAL error on citation if not)
  - check if figureCitations have an HTTP URI (MAJOR error if not)

Common Errors and Possible Solutions

Text

- Word 'XXX' ends a paragraph, but is no respective punctuation mark
- Word 'XXX' starts a paragraph, but is in lower case
- Word 'XXX' appears to be a paragraph, but is not marked accordingly
    - If the lack of punctuation is due to the paper section (eg last word among key-words): click “False Positive or Data Error”
    - Check paragraph structure and/or word-flow. Confer “Set Next Word Relation” and/or “Set Previews Word Relation”.

- Word 'XXX-' appears to be hyphenated, but is not marked accordingly
    - Confer “Set Next Word Relation” and/or “Set Previews Word Relation”.

Treatments

- Treatment 'XXXXX' contains over 30 taxon names
  - Confer if there is no more than one treatment on the same annotation
  - If there is only one treatment: click “False Positive or Data Error”
  - If there are more than one treatments: Split Treatments and check Structure Treatments

- Blocker: Treatment 'XXXX' has fewer than two paragraphs
- Blocker: Treatment 'XXXX' is lacking a nomenclature section
- Blocker: Treatment 'XXXX' is lacking a taxon
- Blocker: Treatment 'XXXX' has fewer than two sections
  - Check Treatment limits and structure: probably merge treatments and reorganize treatment structure

- Cited taxon 'XXX' has a rank in a different group than the treatment taxon
  - Confer if there is no more than one treatment on the same annotation
  - Edit attribute - TaxonomicName: Correct taxonomic status according to paper previous citation - Add/Set Attribute

Illustrations

- Caption 'TABLE XXX.' is not assigned to a target
  - Mark table area, click “Assign Caption”, and click on the corresponding caption
- Caption 'FIGURE XXX.' is not assigned to a target
  - Click on image area, click “Assign Caption”, and click on the corresponding caption

Material Citation

- Materials citation 'XXXXX' has more than two collectors in its respective attribute
- Materials citation 'XXX' has geo-coordinates, but not as a pair
  - Confer if there is no more than one material citation on the same annotation
  - Split material citation and parse

- Materials citation 'XXX' has geo-coordinates, but not as a pair
  - Confer if there is a mistaken coordinate

Bibliographic Reference

- Reference 'XXXX' has an unmatched bracket in the journal name
  - Parse reference and extent the broken annotation

- Reference 'XXXX' has a numberbracket in the journal name
  - Parse reference
  - Unmark and re-mark annotation
  - Click “False Positive or Data Error”

Taxon Names

- Cited taxon 'XXXX' is lacking authority information
  - Edit attribute - TaxonomicName: Correct authority according to paper previous citation - Add/Set Attribute
  - Find previous citation for the taxon: Copy Annotation Atribute - TaxonomicName

Bibliographic References:

- Bibliography (all CRITICAL errors unless indicated otherwise, all per reference):
  - check if each reference has an author
  - check if each reference has a title
  - check if each reference has a year of publication
  - check if each reference has a document internal ID (added by citation tagging)
  - check if each reference is cited by at least one bibRefCitation (MAJOR error on reference if not)

- Bibliographic Citations:
  - check if each bibRefCitation is linked to a reference (CRITICAL error on citation if not)

Taxonomic Names (all BLOCKER errors on taxon names in treatment nomenclature, CRITICAL errors in treatment reference group, MAJOR errors elsewhere in treatments, and MINOR errors outside treatments):

- check if each taxon name has rank attribute
- check if each taxon name has a taxonomic kingdom assigned to it
- check if each taxon name has proper authority information assigned to it
- check if each taxon name of rank family or below has a taxonomic family assigned to it

Treatments:

- check if boundaries of each treatment coincide with paragraph starts and ends (CRITICAL error if not)
- check if each treatment is structured into subSubSections (BLOCKER error if not)
- check if each treatment completely structured into subSubSections (CRITICAL error if not)
- check if boundaries of each treatment subSubSection coincide with paragraph starts and ends (CRITICAL error if not)
- check if each treatment has a nomenclature subSubSections (BLOCKER error if not)
- check if each treatment has a taxon name inside its nomenclature subSubSections (BLOCKER error if not)
- check if each treatment stating ' nov.' subSubSections has a materials citation with type status holotype (CRITICAL error if not)

Materials Citations:

- check if materials citations have repeated detail elements indicating a missing split (CRITICAL error if so)
  - more than one type status
  - more than one country
  - more than one regions
  - more than one coordinate pair (i.e., two coordinates)
  - more than one elevation
  - more than one collector or determiner name label ('leg.', 'det.', etc.)
- check if coordinates in materials citations are either absent or present as a pair (CRITICAL error if not)