**LA METAGENOMICA SCENDE IN TAVOLA: PROGETTO NewTeCH**

Rossana Capoferri, Anna Pozzi, Nelson Nazzicari °e Graziella Bongioni

*Istituto Sperimentale Italiano “L. Spallanzani”, Rivolta d’Adda (CR)*

*°CREA - Council for Agricultural Research and Analysis of Agricultural Economics*

*Research Centre for Animal Production and Aquaculture, Lodi*

Il consumatore attuale è sempre più attento all’origine delle produzioni alimentari del patrimonio nazionale e i prodotti DOP e IGP, più di tutti, offrono garanzia di autenticità, tradizione, gusto, tipicità, legame col territorio di origine, sicurezza e tracciabilità.

Tra le produzioni italiane certificate DOP e IGP i formaggi hanno un ruolo centrale, rappresentando il 57% della produzione nazionale e il 50% delle esportazioni (XVII Rapporto 2019 Ismea-Qualivita). Parmigiano Reggiano e il Grana Padano, tra i marchi più importanti e conosciuti dentro e fuori i confini italiani, rappresentano quindi un importante volano per l’economia nazionale.

Nel 2017 parte il progetto triennale Newtech, acronimo di “New Technologies for Cheese Production”, che vede coinvolti il CREA-ZA di Lodi, l’Istituto Spallanzani di Rivolta d’Adda, l’Università di Parma con il Dipartimento di Scienza degli alimenti e il CNR di Napoli con l’Istituto di Scienze Applicate e Sistemi Intelligenti.

L’ambizioso progetto prevede lo sviluppo di metodologie avanzate a supporto della valutazione di autenticità delle produzioni DOP nel settore lattiero-caseario nazionale, lo studio sperimentale dell’effetto sull’uso di derivati anidri del latte nella produzione di formaggi industriali, ed infine la messa a punto di un nuovo sistema in spettrometria nel vicino infrarosso (NIR) per valutare le proprietà strutturali dei formaggi.

Nonostante il sistema di controllo insito nella certificazione stessa (vigilanza pubblica, controllo di terza parte e autocontrollo) di alcuni prodotti DOP come il Grana Padano, i tentativi di imitazione permangono, facendo emergere quindi l’interesse a utilizzare le sofisticate tecnologie disponibiliper caratterizzarlo, difendendone così l’autenticità.

L’Istituto Spallanzani è protagonista, in collaborazione con CREA-ZA, della linea di ricerca “Approcci innovativi per la lotta alla contraffazione nelle produzioni DOP del settore caseario (Grana Padano)”.

Il laboratorio di Genetica molecolare dell’Istituto ha voluto indagare la possibilità di utilizzare lo strumento del DNA metabarcoding del cloroplasto della cellula vegetale per caratterizzare l’origine geografica del latte utilizzato per la produzione del Grana Padano rispetto al latte non proveniente dall’areale.

Come definito negli articoli 3 e 4 del Disciplinare di Produzione del Grana Padano, questo formaggio è prodotto utilizzando il latte crudo di bovine da latte che devono essere alimentate esclusivamente con alimenti ammessi dal Disciplinare. Tali alimenti devono essere ottenuti da coltivazioni di produzione aziendale o nell’ambito del territorio di produzione del latte destinato ad essere trasformato in Grana Padano per almeno per il 75% della sostanza secca giornaliera somministrata attraverso la razione. (Fig 1: Areale Grana Padano)

Questi aspetti evidenziano quanto forte debba essere il legame di questo DOP con il territorio e con l’area geografica che lo produce, in quanto fortemente caratterizzante nelle sue proprietà organolettiche rendendolo tipico e apprezzato dal consumatore.

La biologia molecolare, in questo contesto di caratterizzazione della zona geografica di origine, può quindi diventare un valido strumento potendo avvalersi di metodi di identificazione di specie basati sul DNA utilizzando marcatori molecolari. Tra questi ha un ruolo importante il “DNA barcode” del cloroplasto come strumento molecolare per verificare la possibilità di tracciare la provenienza del latte attraverso la sua componente vegetale di derivazione alimentare.

“DNA barcode” è il termine coniato per definire una specifica regione di DNA in grado di discriminare le specie, siano esse vegetali o animali, anche partendo da piccoli frammenti: la sequenza genetica identificante può essere infatti vista come un codice a barre che identifica un prodotto, dove a ciascuna barra del codice corrisponde una base nucleotidica, essendo queste 4 (A,T,C,G), le loro innumerevoli combinazioni e quindi sequenze di ripetizioni determinano la diversità tra le specie (fig 2 DNA barcode).

**DNA barcode** Questo approccio di indagine molecolare è stato proposto nel 2003 dal genetista di popolazione Paul Hebert dell’Università di Guelph in Canada. Hebert utilizzò la sequenza del gene mitocondriale codificante per il citocromo c ossidasi 1 **(CO1)** per l'identificazione di identità biologiche. Nel corso del tempo il gene CO1 è stato usato con successo per lo studio della biodiversità nel regno animale, dimostrando unaltissima variabilità interspecie e una bassissima variabilità intraspecie ; vengono gettate le basi per il progetto Barcoding Life.

Nei vegetali Il genoma mitocondriale è diverso, per motivi evolutivi, da quello animale e questo lo rende non adatto come barcode. Per le specie vegetali si sono quindi dovuti identificare altri marcatori, identificandoli nel genoma del cloroplasto: il gene matK e il gene rbcL. (figura A )

La tecnica del DNA barcode in quanto tale non è però sufficiente per rispondere alla necessità di mappare più specie contestualmente presenti all’interno di uno stesso campione in quanto è in grado di identificare una sola specie alla volta. Per un’indagine più profonda è invece necessario avvalersi di un sistema in grado di fare un’identificazione multispecie utilizzando tutto il DNA, anche frammentato, presente in matrici complesse: il DNA metabarcoding.

**DNA metabarcoding** Il metabarcoding è un metodo di valutazione della biodiversità che associa le recenti tecniche di sequenziamento massivo e profondo (next generation sequencing), con la tassonomia basata sul DNA (DNA barcode). I principali vantaggi della metagenomica sono dati dall’ identificazione di migliaia di specie in una singola analisi e dalla possibilità di risalire all’abbondanza relativa delle diverse popolazioni, che possono quindi definire differenti areali.

In questo studio la matrice complessa indagata è il latte, nel quale oltre al DNA bovino, derivato dalle cellule somatiche, residuano in proporzioni non definibili anche il DNA appartenente alle specie vegetali presenti nell’alimento degli animali, che è quello che si vuole caratterizzare, ma probabilmente anche del DNA di diversa natura e di origine sconosciuta.

Al fine di avere una prima mappatura dell'areale del latte del Grana Padano e valutare se le specie vegetali identificate siano sufficientemente rappresentative e discriminatorie tra areali geografici diversi (ad esempio grana e non grana) nell’ambito del progetto sono stati campionati tutti i caseifici dell’areale.

La raccolta dei campioni di latte da parte dei tecnici del Consorzio Grana Padano ha riguardato 136 caseifici della zona di produzione del Grana Padano DOP. Tale attività è iniziata ad ottobre 2017 ed è terminata nel mese di ottobre 2018 con una media mensile di 130 campioni. Tutti i campioni pervenuti in Istituto sono stati codificati e memorizzati in un apposito database per la gestione della biobanca.( fotografia campioni latte in lab)

La Tabella 1 riporta tutti i campioni raccolti e che sono andati a costituire la biobanca, allestita presso l’ Istituto Spallanzani.

Per l’analisi in metabarcoding sono stati scelti dei campioni che fossero il più possibile rappresentativi di tutto l’areale di produzione: i campioni identificati provengono da 13 caseifici (10% dei totali) distribuiti nelle 5 Regioni (Piemonte, Lombardia, Emilia Romagna, Trentino e Veneto) e nelle diverse Province (Cuneo, Brescia, Piacenza, Trento, Verona e Vicenza). (immagine dia)

Per ciascuno dei caseifici individuati sono stati testati 3 campioni sulla base della stagionalità per un totale di 39. Oltre ai campioni del Grana Padano, sono stati aggiunti dei campioni controllo per il latte che provenissero da zone completamente esterne all’areale (Napoli, Sicilia, Puglia e latte di montagna) e campioni di controllo di cui fosse nota la componente vegetale (Soia e due mangimi ammessi dal Disciplinare).

Nel primo periodo di attività l’attenzione è stata tutta incentrata sulla messa a punto di un metodo estrattivo che meglio rispondesse alle esigenze di indagine successive, quindi attraverso un’ampia ricerca bibliografica e diverse prove di laboratorio la scelta è ricaduta su una metodologia di isolamento del DNA (Ponzoni et al, XXXX) con metodi prettamente chimici e senza avvalerci di kit commerciali. La metodica permette l’isolamento di tutto il DNA presente nel latte, indipendentemente dalla sua origine (bovino, vegetale e ignoto) e dalla sua dimensione.

Essendo l’obiettivo di tale studio la caratterizzazione dell’origine geografica del latte attraverso differenze nella componente vegetale, la porzione di DNA specifica (target) indagata tramite Metabarcoding è il gene *rbcL* del cloroplasto, che codifica per la subunità grande dell’enzima ribulosio-1,5- difosfato carbossilasi/ossigenasi (RuBisCO), il più importante accettore di carbonio in tutti gli eucarioti fotosintetici. (fig 3).

La scelta dell’utilizzo di tale marcatore è dettata dal fatto che è un gene altamente conservato all’interno della stessa specie, ma allo stesso tempo molto polimorfico, quindi molto variabile tra specie diverse

Il DNA vegetale che residua nel latte e derivante dalla dieta è fortemente degradato e frammentato a causa del passaggio attraverso il sistema digerente Per verificare quindi che il DNA totale estratto contenga anche il DNA vegetale i campioni sono stati amplificati con primer specifici per il gene RuBisCo (rbcL) del cloroplasto di Phipps et al. (2003): RUB F2 TTGGCAGCATTCCGAGTAAC e RUB R2 GTGAGGCGGACCTTGGAAAG. Tutti i campioni sono stati amplificati ed il frammento atteso di 351 bp è stato visualizzato mediante corsa elettroforetica Alcuni COSA? FRAMMENTI? CAMPIONI? sono stati anche sequenziati COME? SANGER?per confermare l’appartenenza del frammento al gene RuBisCo.

Una volta verificata l’amplificazione di questo frammento i campioni sono stati analizzati con un sistema di Next Generation Sequencing (NGS) e Metabarcoding che prevedono il sequenziamento mirato per porzioni specifiche (target) anche di piccole dimensioni, al fine di ottenere informazioni tassonomiche sulle specie rilevate. Tutti i campioni di DNA (39 campioni estratti dal latte e 11 controlli) sono stati quindi processati nei laboratori dell’IGA Tech di Udine, che si avvale di strategie innovative e di tecnologie di sequenziamento avanzato per tracciare l’origine degli alimenti. Il flusso di lavoro per il metabarcoding del tag rbcl è suddiviso in 4 step: preparazione delle librerie, sequenziamento, analisi dei dati e risultati finali.

Le librerie possono essere preparate con i kit sviluppati per la preparazione delle librerie Nextera. Vengono quindi utilizzati primer specifici per la regione rbcl variabile del gene RuBisCO al fine di caratterizzare la composizione delle comunità vegetali presenti nei campioni. I primer utilizzati sono stati i seguenti: rbcL F 5’- ATGTCACCACAAACAGAGACTAAAGC-3’ and rbcL R 5’-GTAAAATCAAGTCCACCRCG -3’ e le librerie sono state quindi sequenziate sullo strumento MiSeq (Illumina, San Diego,CA) utilizzando la modalità 300-bp paired-end.

La fase di sequenziamento include la generazione di cluster, il sequenziamento e l'identificazione delle basi ed è stata effettuate sul sistema MiSeq che sfrutta la tecnologia di sequenziamento mediante sintesi (Sequencing By Synthesis, SBS Illumina), la più utilizzata chimica di sequenziamento di nuova generazione in tutto il mondo. Il metodo è basato su terminatori reversibili che consente il sequenziamento massivo in parallelo di miliardi di frammenti di DNA, rilevando singole basi mentre vengono incorporate in filamenti di DNA in crescita.

MATERIALI E METODI DIVULGATIVO PER ANALISI BIOINFORMATICA

RISULTATI PRELIMINARI

Il sequenziamento di questa regione nel DNA vegetale, isolato dal latte, produce quindi molte sequenze nucleotidiche che, grazie ad una mirata analisi bioinformatica, permettono di identificare le eventuali differenze genetiche, ma soprattutto permettono identificare le specie vegetali presenti nella matrice latte e di produrre quindi una classificazione tassonomica.

Tutte le sequenze prodotte devono essere quindi “confrontate” con sequenze di riferimento depositate nei database per l’identificazione specifica. Nel momento di stesura di questo articolo i dati derivanti dall’analisi di metabarcoding sono stati interamente raccolti e sono in fase di elaborazione.

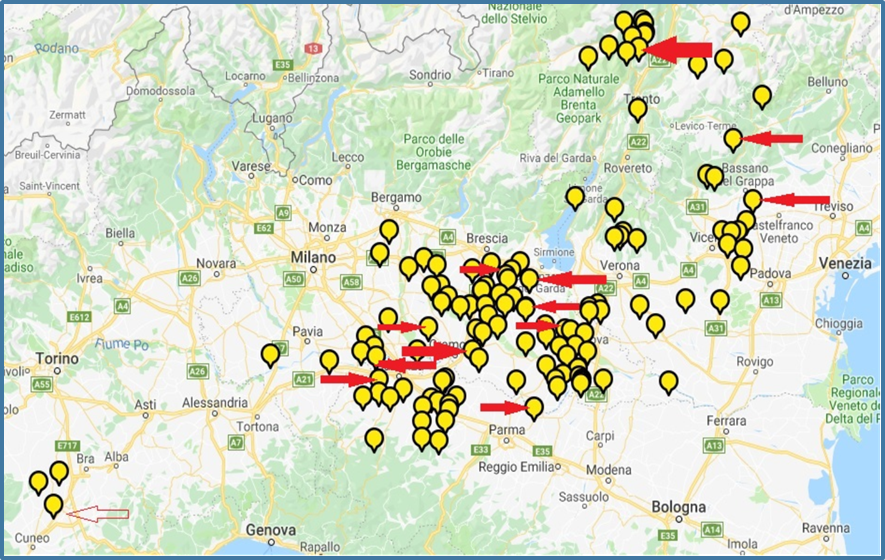
Dalle prime valutazioni dei risultati il metodo estrattivo scelto si è rivelato efficace, sebbene un campione di quelli analizzati non abbia prodotto risultati. Per tutti gli altri campioni, come atteso, è stata riscontrata la presenza in percentuale importante di DNA bovino (circa il 63%). L’analisi bioinformatica del DNA vegetale è ancora in corso e permetterà di valutare la presenza di differenze genetiche in grado di definire una caratterizzazione geografica del latte usato per produrre Grana Padano.

Un primo risultato, del tutto preliminare, è relativo alla metrica di “richness di Shannon-Wiener”. Questo indice di diversità, tra i più usati, misura il numero di specie diverse contenute in ogni campione e le mette in relazione all’effettiva numerosità delle tracce, così che una specie molto rara abbia meno peso di una molto comune. Un primo raffronto, riportato in Figura XXX, mostra come nell’areale del Grana l’indice di richness sia maggiore, ad indicare un’alimentazione più variegata.

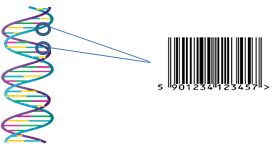
Immagini

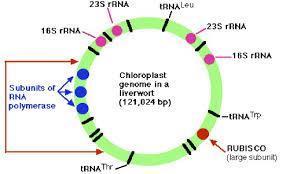
FIGURE

-Fig. 1 Areale Grana Padano

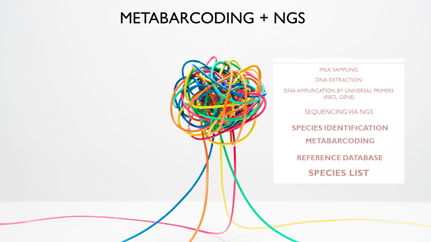


* Fig.2 DNA Barcode





* Fig. 3
* schema processo (da fare meglio)



Milk collection

DNA extraction

bovine

plant

other (e-DNA)

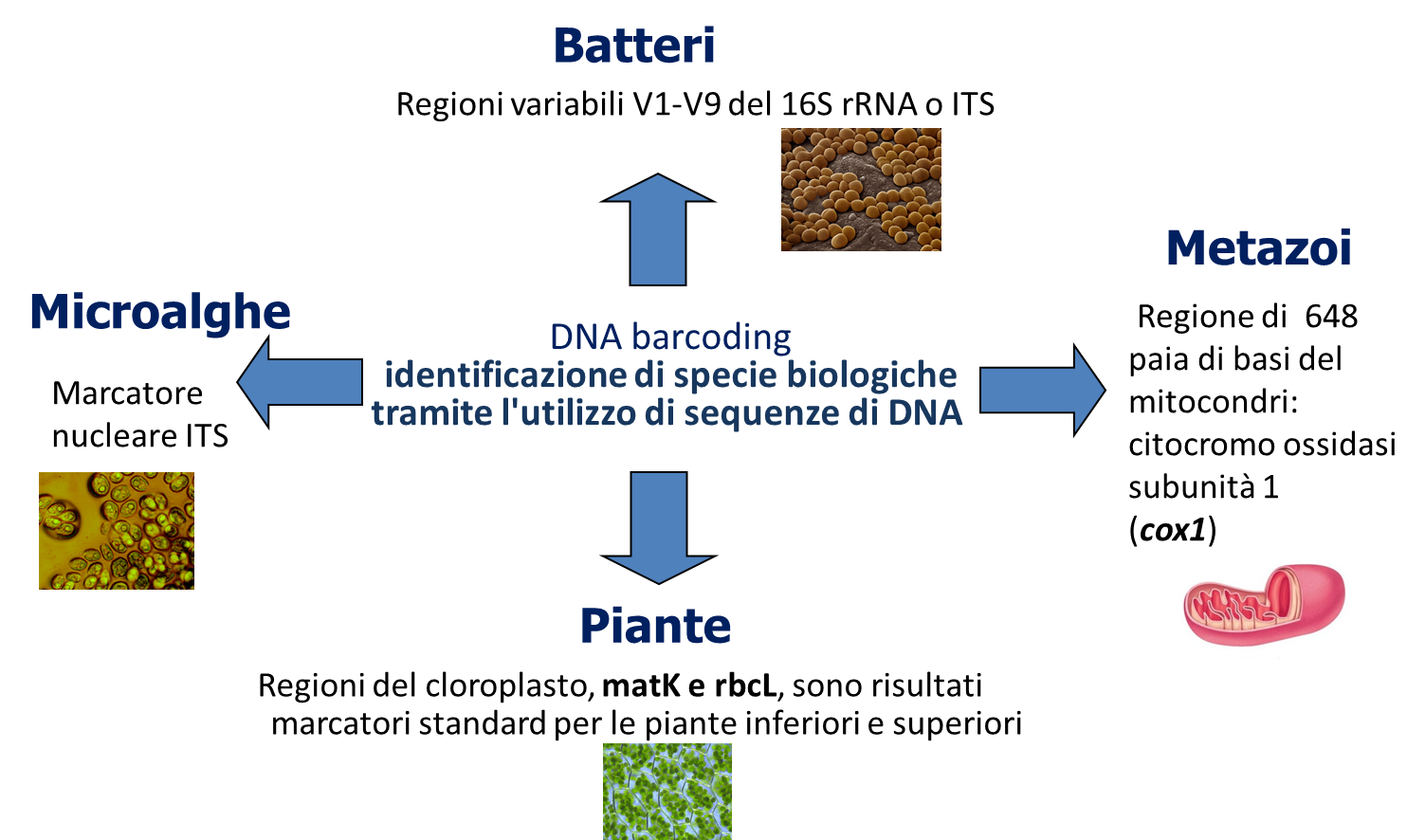
Amplification

NGS

Bioinformatic Analysis

PLANT CHARACTERIZATION

Figura A

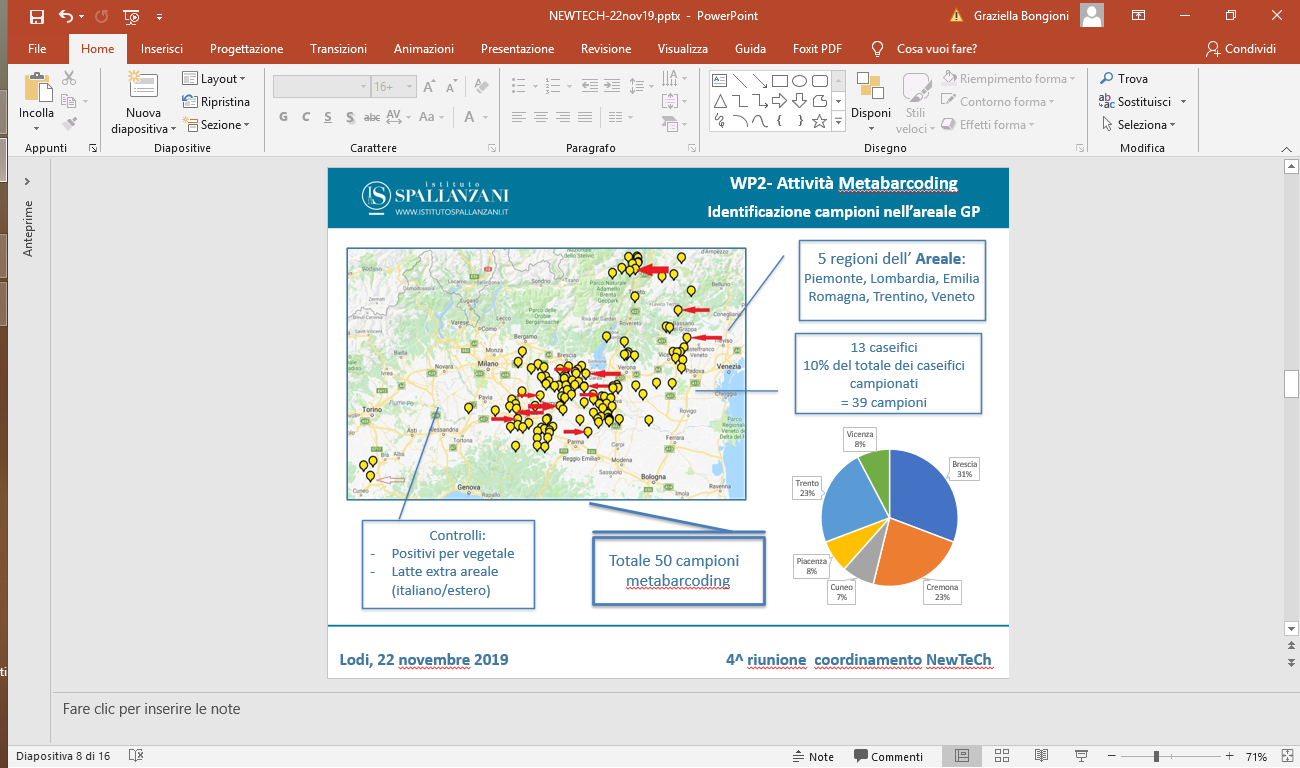






|  |  |
| --- | --- |
| ***Mese*** | ***N° campioni di Latte*** |
| *ottobre 2017* | 15 |
| *novembre 2017* | 133 |
| *dicembre 2017* | 131 |
| *gennaio 2018* | 134 |
| *febbraio 2018* | 136 |
| *marzo 2018* | 135 |
| *aprile 2018* | 132 |
| *maggio 2018* | 134 |
| *giugno 2018* | 131 |
| *luglio 2018* | 126 |
| *agosto 2018* | 115 |
| *settembre 2018* | 120 |
| *ottobre 2018* | 128 |
| ***totale*** | **1570** |

Tab.1 Numero e tipologia di campioni di latte Areale Grana Padano



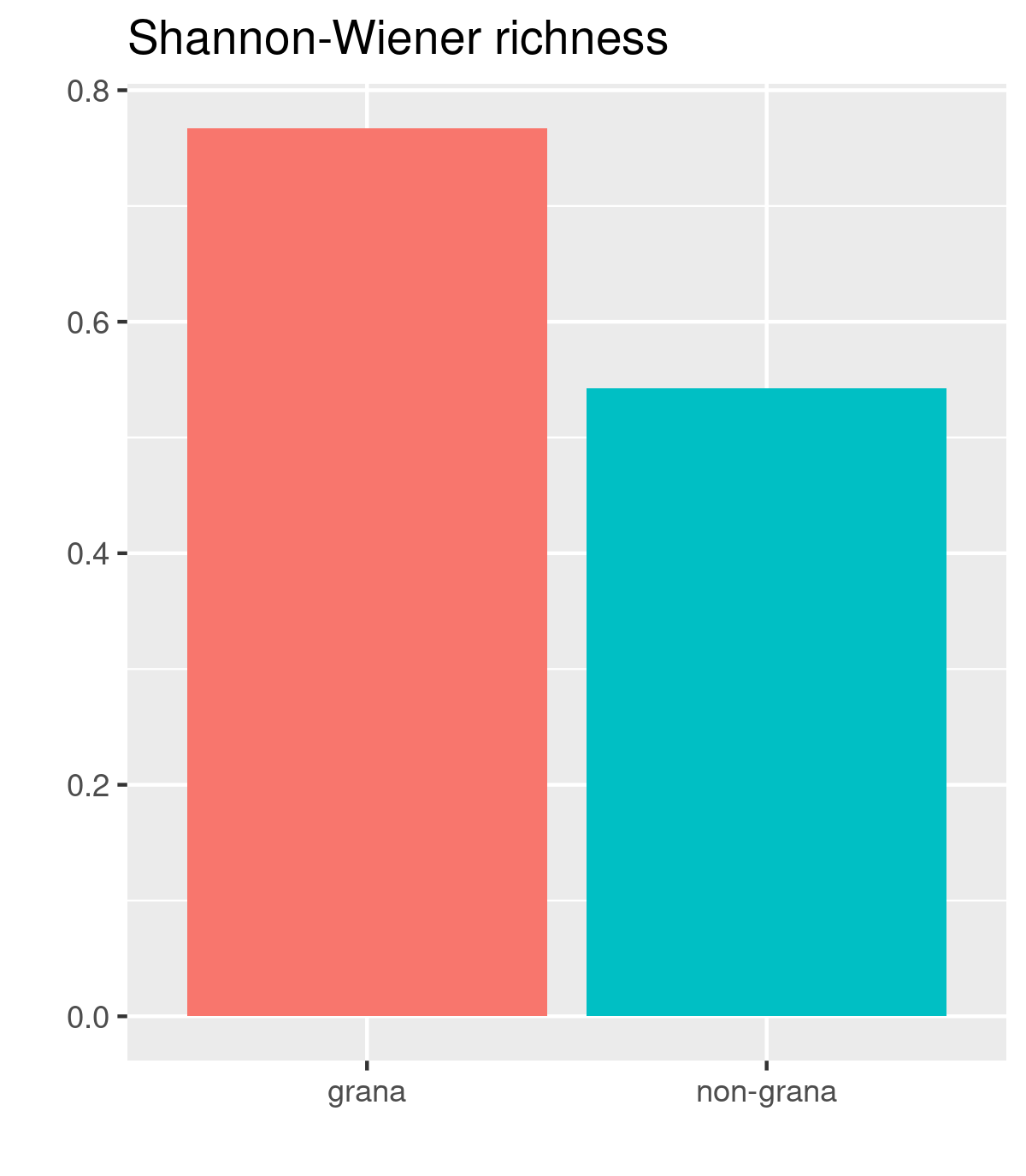


Figura XXX: indice di richness Shannon-Wiener medio, calcolato a livello tassonomico di specie, per i campioni entro e fuori areale